DOCUMENT RESUME

ED 227 168 TM 830 190

TITLE	Bulletin IV: The Technical Manual. The Florida Teacher Certification Examination.
INSTITUTION	Florida State Dept. of Education, Tallahassee. Student Assessment Section.; Florida State Dept. of Education, Tallahassee. Teacher Certification Section.
PUB DATE	81
NOTE	136p.; Some tables may be marginally legible due to small print. For related documents, see ED 193 171-173.
PUB TYPE	Guides - Non-Classroom Use (055) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	Elementary Secondary Education; *Minimum Competencies; *Occupational Tests; State Programs; *Teacher Certification; Teacher Evaluation; *Teaching Skills; Test Bias; *Test Construction; Test Items; *Test Manuals; Test Reliability; Test Validity
IDENTIFIERS	*Florida Teacher Certification Examination

ABSTRACT

The Florida Teacher Certification Examination was developed as a comprehensive written examination for use in initial teacher certification in the state. Applicants must demonstrate mastery of 23 generic minimum competencies in teaching adopted in Florida statutes and State Board of Education rules. Subtests in General Education cover writing, reading and listening, and mathematics. Subtests in Professional Education include physical development, social development, academic development, interpersonal skills, technical skills, and administrative skills. The examination format includes a writing production task, a reading comprehension cloze test, and objectively scored items for other subtests. The development of the examination included the writing and validation of test items within specified competencies and skills. Pilot testing and a final test item review established its content validity. Field testing of the multiple choice items and rater judgments in special pilot testing of the writing test provided reliability evidence. Score scaling, passing score standards, and the bias analysis procedures are described. The appendices include the personnel who contributed to development of the examination, evaluation strategies for competencies and skills, examination coverage, and agendas and materials used in the item review and standard setting panel. (CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

The Florida Teacher Certification Examination
Bulletin IV: The Technical Manual

ED227168

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

C. Ferrell

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 820 190

Copyright State of Florida Department of State 1981

PREFACE

Four bulletins have been developed to provide information about the Florida Teacher Certification Examination. Bulletin I describes the development of the examination and presents the specifications for each of the subtests. It also provides a list of supplemental materials and references and a set of sample items to illustrate the kind of items which will be used in the examination. Bulletin II presents a general description and specifications for each item in the reading, writing, and mathematics subtests. Bulletin III provides specifications for the professional education subtest of the Florida Teacher Certification Examination. It includes a description of the content base and item specifications for each of the professional education items to be included in the examination. This bulletin, the Technical Manual, is the fourth in the series and describes the technical adequacy of the examination including such topics as establishing test reliability, test validity, passing scores, and protecting the test from cultural or ethnic bias.

It is expected that faculty members of teacher education programs and students in these programs will be especially interested in Bulletin I. Directors of Teacher Education Centers and school district staff development directors may also find the information useful. The specific item specifications and other information in Bulletins II and III will probably be of special interest to professionals involved in program development and evaluation. Bulletin IV was designed primarily for measurement professionals.

Please note that the scope of the examination is limited to the essential generic competencies which are assessable by a written examination. There has been no attempt to cover all aspects of teacher training. Many important competencies are assessable only by direct observation, and many competencies are specific to the subject matter taught or the developmental level of the students. It is also important to remember that teacher education is dynamic; it must change to reflect and incorporate new research evidence and wisdom accumulated from experience. For these reasons, even though the examination has been carefully developed and reflects the current state of knowledge and priorities for the general preparation of teachers, the specifications for the examination should not be used as the sole basis for a teacher training program.

The Department of Education encourages professionals to make recommendations to improve the Florida Teacher Certification Examination. Recommendations and other inquiries should be addressed to:

Dr. Garfield Wilson, Administrator
Teacher Certification Section
Florida Department of Education
Tallahassee, Florida 32301

Dr. Thomas Fisher, Administrator
Student Assessment Section
Florida Department of Education
Tallahassee, Florida 32301

TABLE OF CONTENTS

CHAPTER		PAGE
I	DEVELOPMENT OF THE EXAMINATION	1
	Planning	1
	Identification and Validation of the Essential Teacher Competencies	1
	Calendar of Key Events in the Development of the Florida Teacher Certification Examination	2
	General Planning for Examination Development	3
	Development of Subskills for Each Competency	5
	Development of Item and Test Specifications	6
	Writing and Validating Test Items	8
	Creating the Test Items	8
	Pilot Testing the Items	10
	Reviewing the Test Items	11
	Field Testing of the Certification Examination Items	11
	Setting Passing Scores	12
	Preparing for Test Assembly, Administration, and Scoring	12
II	VALIDITY	13
	Development of the Items	14
	Final Test Item Review	15
	Content Review	15
	Bias Review	17
	Technical Review	17
	Analysis of Field Test Data	17
	Summary	17
III	RELIABILITY	19
	Reliability of Multiple Choice Tests	19
	Standard Error of the Passing Score	20
	Reliability Coefficients	21
	Reliability of Scoring of the Writing Test	22
	Summary	23
IV	SCORE SCALING AND PASSING SCORES	25
	Score Scaling	25
	Setting Passing Score Standards	26
	Requirement for Perfection	27
	The Use of Normative Data	27
	Opinion of Judges	27
	Procedures Used for the Florida Teacher Certification Examination	28
	Recommendations	30

V BIAS ANALYSES 33

Developmental Procedures 33
Statistical Procedures 34
TID Comparisons (Angoff) 35
Rasch Analyses 36
Correlational Analysis 37
Summary 37

VI APPENDICES

APPENDIX A - Professional Contributions (Personnel) 39

APPENDIX B - Evaluation Strategies for Competencies and Subskills 49

APPENDIX C - Coverage of the Examination 69

APPENDIX D - Agendas 77

APPENDIX E - Forms and Materials--Item Reviews 87

APPENDIX F - Technical Materials--Standard-Setting Panel . . . 111

APPENDIX G - References 131

LIST OF TABLES

TABLE		PAGE
1.1	Calendar of Key Events in the Development of the Florida Teacher Certification Examination	2
1.2	Participants in Field Test	12
2.1	Item and Person Assignment for Expert Review	16
3.1	Standard Error for Field Test Forms of the Florida Teacher Certification Examination Expressed as Logits . .	20
3.2	Coefficient Alpha for Various Forms of the Florida Teacher Certification Examination Field Test	21
3.3	Estimated Reliability for Lengthened Subtests of the Florida Teacher Certification Examination	21
3.4	Inter-Rater Reliability of Writing Subtest of the Florida Teacher Certification Examination	22
3.5	Coefficient Alpha for Inter-Rater Reliabilities for Writing Subtest of the Florida Teacher Certification Examination	23
5.1	Correlations of Item Difficulties Between Various Groups . . .	38

CHAPTER I

DEVELOPMENT OF THE EXAMINATION

Section 231.17 F.S. requires that, beginning July 1, 1980, each applicant for initial teacher certification shall demonstrate on a comprehensive written examination mastery of the minimum competencies adopted in Florida Statutes and State Board of Education Rules. In order to develop the Florida Teacher Certification Examination (FTCE), the following tasks had to be accomplished: (1) planning, (2) writing and validating test items, (3) field testing the certification examination items, (4) setting passing scores, and (5) preparing for test assembly, administration, and scoring. The procedures by which each task was addressed are described in this chapter.

The events associated with each developmental task are described in chronological order for a given task; however, since many of the tasks overlapped, there is some duplication of time lines across tasks. There is also some redundancy in these descriptions, because a single activity was sometimes used to accomplish more than one task. The calendar of key events is provided in Table 1.1. Reference to this calendar will make clear the overall chronology of the development of the examination.

Planning

The subtasks involved in planning for the examination were (1) identification and validation of the essential teacher competencies, (2) general planning for examination development, (3) development of subskills for each competency, and (4) development of test and item specifications. Each of these activities is described in this section.

Identification and Validation of the Essential Teacher Competencies

The single most important step in the development of an examination is the determination of what it is to cover. From 1975 through October 1980, the Council on Teacher Education (COTE) was involved in the development and implementation of a "competency based model" of teacher certification, which was enacted into law as Committee Substitute for Senate Bill 549 (CSSB 549) and became Section 231.17 F.S. During this time, COTE served as a statutory advisory council appointed by the State Board of Education to advise the Commissioner of Education on all matters dealing with teacher education and certification.[1]

The "minimum essential competencies," which were to be the basis of teacher certification in Florida, were identified in a study conducted by COTE, beginning in March 1975. The first phase of the study was the identification of a set of competencies deemed essential for all certified personnel. Following an intensive literature search, the Council prepared a survey instrument listing 48 competencies and submitted it to a five percent random sample of all certified educational personnel employed in Florida for 1975-76. The survey revealed an extremely high degree of consensus on 23 of the 48 competencies. The 23 competencies identified in the survey were recommended to the Commissioner of Education for incorporation into State Board of Education Rules governing teacher education and certification.

---

[1]The Council on Teacher Education was replaced in 1980 by the Florida Educational Standards Commission.

TABLE 1.1

CALENDAR OF KEY EVENTS IN THE
DEVELOPMENT OF THE FLORIDA TEACHER CERTIFICATION EXAMINATION

Identification and Validation of the Essential Competencies	1975-1976
Development of the Competency-Based Model of Teacher Certification	1977-1978
SBE Rule, Requiring Essential Competencies for Teacher Education Programs	March 1978
Section 231.17 F.S. Enacted	June 1978
Ad Hoc Professional Task Force Established	July 1978
Development and Validation of Subskills for Each Competency	July-October 1978
Task Force Report to COTE	November 1978
COTE Recommendations on Test Development by Contractors	January 1979
Development of Item Specifications	March-June 1979
Review of Item Specifications by Task Force	July 1979
Refinement of Examination Specifications by Measurement Consultants	August 1979
Development of Items by Contractors	September 1979-January 1980
Expert Review of Items	January 27-29, 1980
Field Test of Items	April 17-18, 1980
Analysis of Field Test Data	April-June 1980
Setting Passing Scores	June 24-26, 1980
Approval of Contract for Formatting, Administration and Scoring of Examination	July 15, 1980
Adoption of SBE Rules for the Florida Teacher Certification Examination	July 30, 1980

After extensive publicity and public discussion, the 23 Essential Generic Competencies were incorporated in March 1978, into State Board of Education Rules governing the curricular requirements at institutions of higher education with approved teacher education programs. All of the COTE deliberations, decisions, and recommendations, along with the results of the numerous surveys, are documented in the COTE Minutes for 1975 through 1978. The competencies are listed in Appendix B.

General Planning for Examination Development

In June 1978, immediately after the Governor had signed CSSB 549 into law, a meeting was called by the Department of Education to begin planning for the written examination required by the law. Attending this meeting were members of the COTE Executive Committee, representatives from the two state teachers' organizations and the state administrators' organization, and legislative staff members. This meeting resulted in recommendations to the Department of Education as to timeline, tasks, and professional involvement for the development of the teacher certification examination.

Responsibility for directing the development of the examination was assigned to Dr. Garfield Wilson, Office of Preservice Teacher Education, Department of Education, with COTE serving as the primary advisory group. In July 1978, a Professional Ad Hoc Task Force was convened. The Task Force consisted of six classroom teachers, three teacher educators, one school administrator, and one Department member. The classroom teachers had been recommended for participation by each of the two state teacher organizations. Dr. Jacob Beard, Professor of Education, Florida State University, was selected as a measurement consultant to assist COTE and the Task Force.[2]

The Professional Ad Hoc Task Force was responsible for many activities in the early development of the examination. One of its first duties was to consider the overall design for the examination, an activity conducted with the assistance of Dr. Beard. The Task Force discussed major design tasks and problems and then compiled a set of recommendations which were circulated for review throughout the State. Comments were solicited from all teacher education institutions, school districts, Teacher Education Centers, Department staff, and professional organizations. All reactions and comments were carefully considered by Department staff and the revised recommendations were then presented to the COTE Committee.

At its January 1979, meeting, COTE considered the recommendations developed by the Department. It subsequently adopted the following recommendations concerning the overall design parameters for the teacher examination:

---

[2]The Department also was assisted in the planning phases of this project by Dr. Howard Stoker, Florida State University, Dr. Annie Ward, University of South Florida, and Mr. Robert Feinberg, University of Florida.

1. The test should be an objective-based paper and pencil examination.

2. The test should be organized into two major parts addressing general education and professional education competencies with each part divided into subtests as follows:

   General Education
   - A. Writing
   - B. Reading and Listening
   - C. Mathematics

   Professional Education
   - A. Physical, Social, Academic Development and Interpersonal Skills
   - B. Technical Skills
   - C. Administrative Skills

3. The response format should include:

   - A. Writing--a writing production task that will be rated holistically by selected evaluation experts
   - B. Reading Comprehension--(1) A "Cloze" test
     (2) Reading passages with multiple-choice questions
   - C. Multiple-choice or some other form of objectively scored items for other subtests

4. Specifications for test items should be based on the subskills that are recommended by COTE after consultation with experts. The subskills must be amenable to paper and pencil testing. The 127 subskills for the 23 essential generic competencies and the competencies specified in statutes were identified by an Ad Hoc Task Force in July 1978.

5. Specifications for the test items for each subtest should be developed on a contract basis with Florida Institutions of Higher Education with demonstrable expertise in the area in question.

6. Specifications developed by contractors should be carefully reviewed by COTE and the Task Force.

7. Following approval of specifications by COTE, the State should develop a precalibrated item pool [calibrated through Rasch technology].[3]

8. The item pool should be large enough to allow construction of a new form of the test with each administration for purposes of security.

[3]Minutes of the Council on Teacher Education, January 1979.

9. The items in the precalibrated item pool should be included on appropriate subtests in rough proportion to the median weights assigned by the Task Force.

10. The test will be a requirement for eligibility for a Florida teaching certificate; it will not be a requirement for a baccalaureate or other degree.

11. The test fee should be sufficient to cover the cost of maintenance and administration.

12. The test should be administered at least twice a year at selected locations statewide.

13. Only pass/fail scores shall be reported on each subtest. All subtests must be passed; only subtests failed must be retaken for eligibility for regular certification.

14. The procedure for establishing cut-off scores should be developed by measurement experts, advertised, and incorporated into State Board of Education Rules no later than April 1979.

After much discussion by COTE, the Department, and various consultants, a decision was made to issue a Request for Proposals (RFP) to the various universities in the state for the development of item specifications and items for the examination. It was also decided to issue an RFP to universities and commercial testing agencies for the operational tasks required for the formatting, printing, field testing, administering, and scoring of the examination.

Development of Subskills for Each Competency

At the first Task Force meeting in July 1978, one assignment was to develop subskills for 22 generic competencies (competencies #6 and #19 of the original 23 essential competencies were combined by legislative action). During an intensive week of writing, review, and revision, the Task Force developed 127 subskills as indicators for the essential competencies.

In August, 1978, the subskills were submitted to a random sample of certified personnel in the State to determine how well the subskills were related to the original competencies. All but 12 of the subskills met the criterion for acceptance. Two of the subskills (5d and 14b) which did not meet the criterion were retained because of statutory mandates. This left a total of 117 validated subskills. Of these, twenty-six (26) were distributed across the five general education competencies, and the remaining 91 were distributed across the seventeen professional education competencies. (See Appendix B for the list of competencies and validated subskills.)

The competencies and subskills were reviewed by COTE at its September 1978 meeting. COTE was asked to make some recommendations on the most appropriate evaluation procedures for the competencies. COTE recommended that since the Task Force had developed the subskills it would be the most knowledgeable group for developing recommendations on methods of evaluation.

In October 1978, the Task Force was reconvened with measurement consultants Dr. Jacob Beard and Dr. Annie Ward assisting the group. The Task Force in a two-day session, developed recommendations for the most appropriate evaluation procedure for the essential competencies and subskills. These recommendations indicated the competencies and subskills that could be measured most appropriately by each of three procedures: a comprehensive written (generic) examination, a written specialization examination, or a performance evaluation. (See Appendix B.)

The Department of Education staff, with assistance from Dr. Beard, prepared draft overall test design specifications for the examination and the subtests. The proposed specifications were reviewed by COTE at its September and November meetings. The proposed specifications, including the list of validated subskills, were distributed to all colleges and universities with approved teacher education programs and to statewide professional teacher organizations for review and critique. The reactions from these groups were reviewed by the Department and revisions were made as needed.

Development of Item and Test Specifications

In keeping with the January 1979, recommendations of COTE, an RFP was issued, and in March 1979, contracts were awarded for the development of item specifications which would guide the creation of specific test questions. The contracts awarded were as follows:

Writing (Competency 2)

Florida State University
Dr. Gordon Brossell and Dr. James Hoetker
Project Directors

Reading and Listening (Competencies 3 and 4)

Florida State University
Dr. F. J. King
Project Director

Mathematics (Competency 5)

Florida State University
Dr. Tom Denmark
Project Director

Professional Education: Personal, Social, Academic Development (Competencies 6/19, 20, 21, 22, 23)

Florida State University
Dr. Gary Peterson and Dr. Steven Rollin
Project Directors

Technical Skills (Competencies 7, 8, 9, 10, 11, 12, 13, 14)

Florida Atlantic University
Dr. Willard Nelson and Dr. Anna Nelson
Project Directors

Administrative Skills (Competencies 15, 16, 17, 18)

Florida International University
Dr. Colleen Ryan
Project Director

All contractors were required to have their draft item specifications reviewed by experts from other institutions. A sample of the specifications was also reviewed by COTE at the May 1979, meeting. In July 1979, the Task Force was reconvened to review and critique the item specifications. The Task Force also recommended priorities for the inclusion of the subskills for professional education competencies (#6 through #23) on the first and subsequent forms of the written examination. Seventy-two (72) of the subskills were rated either "essential" or "important," while the rest were rated "not necessary."[4]

The Task Force spent three days reviewing and critiquing the item specifications and prioritizing the subskills. Comments and suggestions for improving the item specifications were compiled.

During the process of reviewing the item specifications, the Task Force also considered the general test specifications and further refined them. It was recommended that Competencies 1 (oral communication) and 3 (listening) should be deleted from the written examination but should be assessed by performance evaluation during the year-long internship.

In August 1979, a full day work session was held with DOE staff and measurement consultants Dr. Jacob Beard, Dr. Howard Stoker, and Dr. Annie Ward. The purpose of this session was to review and refine the proposed specifications for the examination. After reviewing the item specifications, Task Force recommendations, and previous COTE recommendations, the measurement consultants made the following tentative recommendations about the make up of the examination:

1. The examination should consist of four subtests: Writing, Reading, Mathematics, and Professional Education.[5]

2. The Writing examination should be 45-60 minutes in length, including time for reading and choosing between topics.

3. The Reading examination should be approximately 30 minutes in length and include a maximum of 100 items in a modified cloze format.

---

[4]Most of those rated "not necessary" were judged to be covered in other subskills.

[5]Three professional education subtests had been recommended by COTE in January 1979.

4. The Mathematics examination should consist of approximately 40 items with one hour of testing time.

5. The Professional Education examination should consist of approximately 100 items with 2½ hours of testing time.

The consultants also recommended that the examination should be administered on Saturdays four times during the first year on the same dates at various locations throughout the state; that the examination should be scheduled for a full day, with mathematics, reading, and writing in the morning and the professional education examination in the afternoon; that scores should be reported on each of the subtests; and that persons scheduled to retake one or more of the subtests should be tested during a regularly scheduled testing administration.

In September 1979, the Department entered into a contract with the University of South Florida to provide psychometric assistance to the examination project. Dr. Annie Ward was assigned this responsibility. She subsequently assisted the Department in the production of the examination and the publications describing the examination. Dr. Ward finalized the specifications for the examination and for each subtest and compiled three bulletins which presented the test and item specifications. These bulletins are: The Florida Teacher Certification Examination Bulletin I: Overview; Bulletin II: The General Education Subtests; and Bulletin III: The Professional Education Subtest.

Writing and Validating Test Items

In order to generate a pool of validated test items it was necessary to: (1) create the test items, (2) pilot test the items, and (3) conduct a review of the items. The procedures used to accomplish each task are described in this section.

Creating the Test Items

The Department negotiated with University personnel in Florida for revising some of the item specifications and for writing items in several areas. The contractors who prepared the item specifications were invited to submit a proposal for writing items according to those specifications. The Technical Skills area (competencies 7, 8, 9, 10, 11, 12, 13, and 14) was divided into two parts. The specifications contractor for that area agreed to prepare items for competencies 7, 12, and 14 and a new contractor was secured for competencies 8, 9, 10, 11, and 13.

Contracts were awarded as follows:

Reading

Florida State University
Dr. F. J. King, Project Director
21 test passages, each containing 10 items

Writing

Florida State University
Dr. Gordon Brossell, Project Director
Validation of topics and evaluation of three presentation modes

Mathematics

Florida State University
Dr. Tom Denmark, Project Director
102 mathematics items

Professional Education

Florida State University
Dr. Gary Peterson, Project Director
100 items for competencies 6, 20, 21, 22, and 23
(Personal, social, and academic development)

Florida Atlantic University
Dr. Willard Nelson and Dr. Anna Nelson
Project Directors
80 items for competencies 7, 12, and 14
(Diagnosis and assessment)

Florida State University
Dr. Walter Wager, Project Director
80 items for competencies 8, 9, 10, 11, and 13
(Teaching and learning)

University of West Florida
Dr. Ronald Peake, Project Director
99 items for competencies 15, 16, 17, and 18
(Administrative/classroom management skills)

In addition to writing items, contractors were required to arrange a series of reviews to establish the relatedness of the items to the validated competencies and to eliminate obvious technical problems. These review processes included:

1. Internal review. All items were reviewed by the contractor's staff before the items were approved.

2. External review. Each item was reviewed by a minimum of four experts. They included classroom teachers, content specialists (either university or public school), and measurement specialists. Each group also included both sexes and representatives of the major ethnic groups.

3. One-on-one administration. Each item was administered individually to at least three persons, with the examinee solving the item aloud. The examiner noted the types of problems examinees had with each item.

After the contractors completed their review and revision of the items, the items were submitted to the Department for review and critique. The items were further revised as needed and prepared for the pilot testing.

Pilot Testing the Items

The contractors conducted a pilot test for their own items, sorting the items into two or three forms for convenience in administration. Common items were provided for the various forms of the pilot tests so that they could be calibrated on a common scale. The sample of students who took the pilot tests was made up of seniors at teacher education institutions of the state. The institutions in which the pilot tests were conducted are indicated below. Each form of each subtest was taken by at least 200 students, except for the Reading test, where the samples were 50 for each form.

Reading	Florida State University Florida A & M University
Writing	Florida State University University of South Florida
Mathematics	Florida State University University of Central Florida Florida A & M University Bethune-Cookman College University of Miami Jacksonville University
Professional Education	(Personal, Social, and Academic Development) Stetson University University of North Florida University of South Florida University of West Florida Bethune-Cookman College (Diagnosis and Assessment) Florida Atlantic University Florida International University Nova University (Teaching and Learning) Florida State University University of Florida (Administrative/Classroom Management Skills) Stetson University University of Florida University of South Florida University of West Florida

Data from the pilot tests were analyzed to secure traditional item statistics. In addition, a contract was negotiated with Florida State University to analyze the data using the Rasch model. Data from both these analyses were available for the final review by experts.

Reviewing the Test Items

A panel of experts was assembled in Tallahassee on January 27, 1980, to review the items. There were three aspects of the review process: content validation, technical qualities, and absence of bias. The review team consisted of classroom teachers, school administrators and supervisors, teacher educators, and measurement experts. The reviewers had been selected in a way that provided expertise in all areas of the examination, as well as expertise in measurement, linguistics, and bias. Three consultants helped to plan and conduct the review process: Dr. William Mehrens, Michigan State University; Dr. Howard Stoker, Florida State University and ETS; Dr. Tom Freijo, formerly at University of South Florida, now President of PDE Associates. The complete roster of participants is included in Appendix A. The review procedures are described in Chapter II, and the agenda is presented in Appendix D.

Field Testing of the Certification Examination Items

The field test of examination items was conducted by the Student Assessment Section, Division of Public Schools, Department of Education. All items which had been accepted by the review panel, or which had been edited, revised, or rewritten according to recommendations of that panel, were assembled into seven field test forms. Several common items were provided for each subtest to be used as linking (anchor) items. The various forms of the test booklets were randomly distributed to the examinees. The field test was conducted on April 17 and 18, 1980, with 1,186 examinees at 14 sites located throughout Florida. The sites were selected to represent public and private schools, schools of varying size, and schools with varying ethnic populations.

Most of the examinees were students who were completing a teacher education program and were completing their internships. In addition, a small group of non-degree Vocational/Industrial teachers who were completing their professional preparation were included. Participating institutions were asked to try to secure total participation of all eligible students. However, at some institutions, the term had ended and many students were no longer available; other institutions had interns scattered over the entire state and communication with some of them was difficult.

During the field test, a few participants were unable to complete all the items. Therefore, the analyses were limited to those examined who completed each subtest. Since all test booklets were arranged so that Reading items were first, then Mathematics, and finally, Professional Education, most of the incomplete papers were in the Professional Education area. The number of cases included in the analyses is shown in Table 1.2.

TABLE 1.2

PARTICIPANTS IN FIELD TEST

	Form Number							
	1	2	3	4	5	6	7	Total
Reading	170	172	168	171	168	167	170	1186
Mathematics	170	172	168	171	168	167	170	1186
Prof. Educ.	164	166	158	164	162	161	166	1141

The Department of Education data center scanned the answer sheets and provided distributions of scores for the total group as well as for the various institutions, program areas, and sex and ethnic groups. In addition, Dr. Jacob Beard at Florida State University provided Rasch analyses, item-total and inter-item correlations, reliability indexes and bias analyses. The analyses are reported in later sections of this manual.

Setting Passing Scores

Passing scores for each subtest were determined by a panel of twenty-one judges, all of whom were either current or past members of COTE and who had been involved in the development of the examination. The panel was made up of eight classroom teachers, five school administrators, five teacher educators, and three community representatives. (See Appendix A for a complete roster.)

The procedure used for the Reading, Mathematics, and Professional Education subtests involved identifying items which represented the level of performance which was acceptable. For the Writing subtest, various levels of prescored writing samples were considered.

As a result of these deliberations, passing score recommendations were made to the Commissioner of Education for each subtest. These recommendations were adopted as a rule by the State Board of Education on July 30, 1980. The procedures are described more fully in Chapter IV of this bulletin and in the minutes of COTE, June, 1980.

Preparing for Test Assembly, Administration, and Scoring

The operational tasks of preparing test forms, administering the tests, and scoring the answer sheets during the 1980-81 school year were conducted by the Office of Instructional Resources (OIR), University of Florida. A contract for these services was awarded through a competitive bidding process. OIR was responsible for assisting the Department conduct the first three test administrations. These administrations, which took place at several different sites across Florida, are described in other publications and reports prepared by the Department of Education.

CHAPTER II

VALIDITY

According to the Standards for Educational and Psychological Tests published by the American Psychological Association (APA, 1974), test validity is concerned with inferences appropriately made from test scores. These inferences may be related to (1) what is being measured by a test and (2) what can be inferred about other behavior.

Jaeger (1979) classified the two types of inferences as those which relate to a well-defined domain of tasks (proximal) and those which relate to some ultimate criteria outside the sampled domain (distal). The former is of importance to the Florida Teacher Certification Examination because of the wording of Section 231.17 F.S.

Section 231.17 F.S. provides that "Beginning July 1, 1980 .... each applicant for initial certification shall demonstrate, on a comprehensive written examination and through such other procedures as may be specified by the state board, mastery of those minimum essential generic and specialization competencies and other criteria as shall be adopted into rules by the state board ...." Note that the statute addresses only the status at time of certification, which is a proximal concern, and does not require that any inferences be made about the distal concern, i.e., whether the test is directly related to the excellence of teaching. The domain of interest was that of entry level skills.

Furthermore, validity, according to Cronbach (1971), relates to whether or not the test items adequately sample the "universe of tasks the developer intended to measure ...." For the Florida Teacher Certification Examination, the broad dimensions of the intended domain were identified in Statute 231.17 as follows:

1. The ability to write in a logical and understandable style with appropriate grammar and sentence structure;
2. The ability to comprehend and interpret a message after listening;
3. The ability to read, comprehend, and interpret orally and in writing, professional and other written materials;
4. The ability to comprehend and work with fundamental mathematical concepts; and
5. The ability to comprehend patterns of physical, social, and academic development in students, and to counsel students concerning their needs in these areas.

The extent to which the Florida Teacher Certification Examination provides data which allow inferences to be made about these dimensions may be determined by examining the adequacy of the set of operations used in test construction. The steps through which the Florida Teacher Certification Examination was developed are described in this chapter.

Development of the Items

The intended coverage of the examination was determined by a process which used professional consensus to (1) identify competencies which should be demonstrated as a condition of certification, and (2) identify subskills associated with each competency. The procedures by which the intended coverage was identified included surveys of the profession, reviews by COTE, reviews by the ad hoc Task Force of COTE, and reviews by teachers and other educational personnel. By these processes, a set of competencies and subskills were developed which defined the domain to be covered by the examination. The domain was further explicated by the various contractors when they prepared the item specifications.

The specifications for each subtest and for items of the Florida Teacher Certification Examination have been published in three bulletins: Bulletin I: Overview; Bulletin II: The General Education Subtests - Reading, Writing, Mathematics; Bulletin III: The Professional Education Subtest. These bulletins were distributed to all Florida teacher education institutions and school system personnel offices in the fall of 1979. An overview of the coverage of the examination is provided in Appendix C.

For the Reading subtest, the domain of materials was derived from materials commonly in use in Florida teacher education institutions. The materials and the procedures used to identify them are described in Bulletin II.

For Mathematics, a survey was made of teachers and other educators to (1) identify subsubskills for the competencies and subskills and (2) determine the relative weights of other dimensions of the test. The specifications derived from these surveys were published in Bulletin II.

For the Writing subtest, it was necessary to determine (1) the kinds of topics to be used and (2) the method of presentation of the topics. A study was conducted to resolve these issues and the results are presented in a report to DOE by Brossell (1980), "Validation of Topics and Comparison of Three Presentation Modes for the Writing Subtest of the Florida Teacher Certification Examination." The specifications for the writing subtest and scoring standards are included in Bulletin II.

For the Professional Education Subtest, contractors identified specific rules, principles, theories, etc., in the professional literature which serve as the basis for both the specifications and the items. This body of supplemental material was combined, analyzed, and organized into the "content base" to which all specifications of the Professional Education subtests have been referenced. Both the specifications and content base were published in Bulletin III.

After the intended coverage was identified and explicated, items were written and pilot tested and then compared once again to the skills being measured. This last item review step included (1) a check that each item addressed the intended competency subskill and specification and (2) a check that each item was free of the invalidating effect of bias or technical problems.

Final Test Item Review

The final item review was carried out in two steps. First, items were reviewed by a special review panel which assembled in Tallahassee on January 27-30, 1980; then the items were field tested with a representative sample of seniors in teacher education programs and reviewed once again by Department of Education staff.

The review panel was comprised of classroom teachers, teacher educators, and administrators. Reviewers were selected to provide coverage of all content areas of the examination and to include women and minorities. (See Appendix A for roster.)

The review process was organized into two tasks. Task I included the following activities:

1. Keying the item, to check the accuracy of keyed response
2. "Blind" Traceback as follows;
   (a) For Mathematics, tracing the item to the subskill and the other specification categories
   (b) For Professional Education, tracing the item to the competency, subskill, and content category(ies)
3. Rating the item for appropriateness

The items were stratified by test area. Reviewers were also stratified by area of expertise. Items were assigned to reviewers in such a way that the following requirements were met:

1. All items were keyed and "traced back" by at least three specialists in the area of content
2. Each item was rated for appropriateness by at least five educators spread across various specialized groups
3. No reviewer was required to trace back more than 60 items
4. Each reviewer also keyed and rated some additional items as to appropriateness

The second task included a review of item content, bias, and technical quality.

Committee assignments were made according to the special qualifications of each reviewer and to the needs for each type of expertise. Review forms were developed to guide the review procedure. All review forms are included in Appendix E.2.

Content Review

The content reviews were conducted by four panels selected to represent specialized areas of the test. Items were also sorted into four groups. The assignment of items and persons to groups is indicated in Table 2.1.

TABLE 2.1

ITEM AND PERSON ASSIGNMENT
FOR EXPERT REVIEW

Test Area	Review Specialists
Mathematics (102 Items)	Mathematics Teachers and Professors Mathematics Education Professors Mathematics Supervisors
Professional Education A. Assessment, diagnosis, measurement	
Competencies/Subskills: 6b; 7 (all); 10h; 14 (all); 18 (all) (approximately 115 items)	Measurement Professors Supervisors of Testing and Evaluation Measurement and Evaluation Consultants Classroom Teachers
Professional Education B. Teaching and Learning, Methods, Materials, Procedures	
Competencies/Subskills: 9 (all); 10b, c, d, f; 11c, d, e; 12 (all); 13 (all); 15 (all); 16 (all); 17a, b, e, f; 20d (approximately 130 items)	Curriculum Methods Professors Curriculum Supervisors School Administrators Classroom Teachers
Professional Education C. Personal, Social, Academic Development	
Competencies/Subskills: 6a, d; 10a, e; 11b; 17c, d; 20b, c; 21b, c, d; 22d; 23b (approximately 124 items)	Education Psychology/Guidance Professors Guidance Counselors, School Psychologists Pupil Personnel Supervisors Classroom Teachers

Each panel member was given a specially designed review form on which to record his or her comments. Separate review forms were provided for the Mathematics and Professional Education subtests. The forms required that the items be rated as to adequacy, relevancy, behavior required, and general item characteristics.

Bias Review

The panel to review the items for potential bias was composed of minority persons, women, and experts trained in linguistics. Many of these people were experienced item reviewers; all were knowledgeable about education. Review forms were prepared which required the reviewers to determine that each item (or passage, in the case of the Reading subtest) was free of the appearance of bias and of biasing elements. (See Chapter V for additional information on the bias review.)

Technical Review

Panelists for the technical review included both measurement and language arts experts. Both groups examined the actual items; in addition, the measurement experts considered item data from the pilot tests. Commonly accepted guidelines for multiple-choice items and review forms embodying the guidelines were provided to each panelist. Experts also noted any technical defect not covered on the form provided.

Each aspect of the Task II reviews was conducted independently. As the reviews were completed for each subtest area, the project staff collated the Task I reviews and the three Task II reviews and made a decision as to the next step for each item. Possibilities included (a) accept the item as is, (b) refer the item to the re-write team, and (c) return the item to the contractor for replacement.

The third task was the rewriting and re-review of items. Some reviewers were assigned to rewrite teams to work with the contractors to revise the items as needed. Contractors were responsible for producing items in a final form which were acceptable to the reviewers and to the Department of Education.

Analysis of Field Test Data

The final step in validating the items was the review of field test data. (See Chapter I for a description of the field test procedures.) Items which did not perform well (i.e., which did not "fit" the Rasch Model; did not correlate well with other items, or which did not discriminate well) were removed from the item bank. These items were either thrown out permanently or revised and subjected to another field test. Staff of the Department performed these tasks during the summer and fall of 1980.

Summary

The validity of the Florida Teacher Certification Examination has been well established. The validity rests primarily upon the content validity of the test. No claim is made that the test scores will have predictive validity--i.e., be able to predict the success of a prospective teacher in an actual classroom situation following passing the examination. It only is claimed that the tests adequately measure the skills for which they were designed. This validity rests upon the manner in which the tests were designed and the care and detail with which the content area domains and test items have been described.

CHAPTER III

RELIABILTY

Reliabilty may be defined as the degree of consistency between two measures of the same thing. The reliability considerations for the Florida Teacher Certification Examination multiple choice subtests (Reading, Mathematics, and Professional Education) are different than those of the Writing test, a production writing sample. With the multiple choice tests the primary concern is for the reliability of an individual's score, while for the Writing test, the concern is for the reliability of the judges' ratings. The data for the reliability studies for the multiple-choice subtests were obtained from the field test conducted in April, 1980. (See Chapter I for a description of field test procedures and N's for each form.) Reliability data for the Writing subtests were obtained in a study conducted for the Florida Department of Education by Brossell (1980), "Validation of Topics and Comparison of Three Presentation Modes for the Writing Subtest of the Florida Teacher Certification Examination."

Reliability of Multiple Choice Tests

An individual's score on a test is made up of the "true score" and an "error score." The true score may be thought of as the "domain" score. Since it is not feasible to actually give an individual all items of a given domain, the domain is sampled to prepare alternate forms of a test. If an individual took several forms of a test, all constructed by sampling from the defined domain of items, scores on the various forms of the test would not vary except as a result of random errors. Such random errors come from item sampling errors and from changes in the individual from one test to another--attention, interest, fatigue, etc.

Standards for Educational and Psychological Tests (APA, 1974) states that ".... 'reliability coefficient' is a generic term. It can be based on various types of evidence; each type of evidence suggests a different meaning." Reliability evidence is generally of two types:

1. internal consistency, which is very important if items are viewed as a sample from a relatively homogeneous universe.

2. consistency over time, which is very important for tests which are to be used for repeated measurement. However, the possibility that scores will change as a result of developmental or educational influences must be considered.

For the FTCE the important reliability issue is the extent to which an individual's performance level would be judged the same regardless of which form of the test is taken. This is an internal consistency question since applicants will not ordinarily take the examination more than one time unless they have failed it. In that case, they may have taken additional preparation directed at improving their performance, so a difference would be expected between scores made from one time to another. However, it is essential that an individual's score at one point in time be consistent with his or her score on another form of the test taken at the same time or with no intervening learning. That is, an

individual's performance can be generalized to his or her performance on the domain covered by the test. This type of reliability is assessed by measures of internal consistency. Two types of reliability estimations are reported for the field test data. First, the standard errors of scores at the passing score level of each subtest are examined, and second, the coefficient alpha is reported for each form of each subtest.

Standard Error of the Passing Score

The standard error of a test score may be used to identify limits that have a defined probability of including the true score. For example, the probability is approximately 68% that an individual's true score lies within the range of the observed score ±1 standard error, and the probability is approximately 95% that the true score lies within the range of ±2 standard error from the observed score. A reliable test will have small standard error values.

For the FTCE the decision point is at the passing score level, as established by State Board of Education Rules. Therefore, it is essential that the examination be very reliable at the passing score point on each subtest. Table 3.1 reports the standard error for each field test form of each subtest at the passing score point.[6]

TABLE 3.1

STANDARD ERROR FOR FIELD TEST
FORMS OF THE FTCE EXPRESSED AS LOGITS

	Field Test Length	Passing Score	Form Number							Median
			1	2	3	4	5	6	7	
Reading	50	1.40	.36	.36	.37	.36	.35	.36	.36	.36
Math.	27-28	1.00	.46	.47	.45	.46	.45	.47	.46	.46
Prof. Ed.	69-70	.25	.27	.27	.26	.27	.27	.26	.27	.27

These standard errors are quite acceptable. Furthermore, since the field test forms for every subtest were much shorter than they will be on the actual examination and since some items were included on the field test which did not meet acceptable standards for inclusion in the actual examination, these standard errors are expected to be smaller for the actual administration.

---

[6]Passing scores and Standard Errors are reported as logits. For a definition of logit see Chapter IV.

Reliabilty Coefficients

A second reliability consideration is that a score on one form of the examination should be comparable to scores on other forms of the examination. Part of this concern is handled by careful calibration of items and person scores so that the various forms are equated. (See Chapter IV.)

Another way of looking at reliability is to consider the reliability coefficients for random collections of items from the item pool, such as those provided by the seven field test forms of the examination.

Coefficient alpha for each of the field test forms of the various subtests is reported in Table 3.2. These coefficients are remarkably consistent across forms and are acceptably high.

TABLE 3.2

COEFFICIENT ALPHA FOR VARIOUS FORMS OF THE FTCE FIELD TEST

	No. of Items	Test Form							Median Alpha
		1	2	3	4	5	6	7	
Reading	50	.82	.81	.86	.82	.83	.77	.83	.82
Math.	27-28	.83	.84	.81	.86	.88	.82	.83	.83
Prof. Ed.	69-70	.81	.79	.79	.80	.82	.82	.82	.81

Because the field test forms of all subtests were shorter than the actual forms of the examination will be, the Spearman Brown prophecy formula can be used to estimate the reliability for tests of the length of each subtest of the actual examination. Estimations of the total test reliability derived through this process are shown in Table 3.3. Again, these values are quite acceptable.

TABLE 3.3

ESTIMATED RELIABILITY FOR LENGTHENED SUBTESTS OF FTCE

	Number of Items	Proposed Length (L)	Median Observed Reliability	Estimated Reliability
Reading	50	80	.82	.879
Math.	27-28	40	.83	.883
Prof. Ed.	69-70	100	.81	.859

Both corrected and uncorrected reliabilities meet the usual standards for educational achievement tests.

Reliability of Scoring of the Writing Subtest

The major reliability concern for the Writing test is the inter-judge reliability of the ratings. In the study from which the reliability data were obtained, essays were written by 360 teacher education students at two teacher universities.[7] Each essay addressed one of six topics. The judges were selected and trained by the same procedures which are being used in the actual administration of the examination. Each essay was rated independently by three judges with a referee to reconcile discrepant scores. Two approaches were used to estimate the reliability. Each of these is described below.

First, four indexes of inter-rater agreement were computed and compared with "target" values derived from the literature. These are reported in Table 3.4.

TABLE 3.4

INTER-RATER RELIABILITY OF WRITING SUBTEST OF FTCE

	Raters' Level	Target Level
Index 1 -- % Complete Agreement	32.2	30-40
Index 2 -- Average % Two of Three Raters Agreeing	98.3	80-90
Index 3 -- Average % Agreement by Pairs as to Pass/Fail	81.3	80-90
Index 4 -- % Complete Agreement about Pass/Fail	71.7	70-80

On three of the four indexes the rating team's level of consistency fell within the target ranges; in one case, Index 2, the team's level exceeded not only the target range, but also the level it had achieved in the training session.

In addition to the four indexes, Coefficient alpha was computed for pairs of raters and for the rating team both before and after the substitution of the referee's ratings. This coefficient indicates the expected correlation between the ratings of the team on this task and those of a hypothetical team of similarly comprised and similarly trained raters doing the same task. Table 3.5 presents the coefficients.

---

[7]The reliability of the scoring of the writing test also was monitored during the scoring process at the University of Florida for each actual test administration. The results of these studies are reported elsewhere.

TABLE 3.5

COEFFICIENT ALPHA FOR INTER-RATER RELIABILITIES
FOR WRITING SUBTEST OF FTCE

	Without Referee's Ratings	With Referee's Ratings
Raters 1 and 2	.619	.640
Raters 1 and 3	.720	.799
Raters 2 and 3	.686	.815
Raters 1, 2, and 3	.759	.828

The figures in Table 3.5 reflect the effect of the referee's ratings on the team's between-rater consistency, increasing the level of reliability in every instance and increasing it substantially in some. The most important coefficient --that of raters 1, 2, and 3 (i.e., the whole team) with the referee's ratings-- is, as would be expected, the highest, since the reliability of a group of trained raters generally increases as its number increases and since the substitution of a referee's ratings is, in and of itself, a deliberate upward adjustment in inter-rater reliability. The level of reliability achieved by the rating team meets acceptable standards for such ratings.

Summary

The Florida Teacher Certification Examination has adequate reliability for the purposes for which it was designed. The reliability of the multiple choice tests was calculated from data derived from the April 1980 field test. Reliability coefficients were expressed in terms of standard error, coefficient alpha, and Spearman-Brown Prophecy values. The test reliability is further monitored each time the test is administered, but these data are reported elsewhere.

For the writing test, reliability of raters' judgements was monitored in a special pilot test of the procedures. The raters' decisions during the scoring of the actual tests is monitored by the test support contractor, the University of Florida, and is reported elsewhere.

CHAPTER IV

SCORE SCALING AND PASSING SCORES

In order to set passing scores on the multiple choice subtests of the FTCE, there were two major tasks to be accomplished. First, it was necessary to determine the performance level on each subtest which would be judged satisfactory and secondly, it was necessary to develop a system to equate performance across different forms of the examination. Although these are separate tasks, it was considered desirable that the outcome of both be incorporated in a single rule; that is, the rule for passing scores on the examination would be expressed in terms of a scaled score which would be used to equate various forms of the examination.

Score Scaling

The matter of the scaled score was addressed first. Very early in the development of the examination it was decided to use a non-normative scale, and the Rasch model was selected as the scaling and equating method to be used.[8]

The Rasch model is one of a family of measurement models developed by Georg Rasch. It is now being widely used in large-scale testing programs. Both item difficulty and person ability are calibrated on the same scale. Person ability is expressed as a function of the difficulty level of items which an examinee can answer correctly and item difficulty is expressed as a function of the ability of examinees who answer the item correctly.

Both person ability and item difficulty are expressed as a logit, a term coined from "logarithm odds transformation." The logit for a person's ability is the natural log odds for succeeding on items selected to define the "zero point" (.00) on the scale. The logit for an item's difficulty is the natural log odds for eliciting failure from persons at the "zero point" (.00) on the scale.[9]

Logits range from approximately -6.00 through .00 to +6.00, with scales for most tests ranging from -3 to +3. The logit scale can be used to interpret item difficulty and person ability. However, the logit scale has two drawbacks. First, it is a decimal scale, and second, it has negative numbers. In order to make the scale more convenient to use, it should be transformed to cover a different range of values. In making the transformation, several properties for the new scale were important. First, the transformation should be mathematically simple. Second, the transformed scale should preserve all the information in the logit scale. Third, the scale should not be easily confused with other scales in common usage.

---

[8]Minutes of the Council on Teacher Education, January 1979.

[9]For additional information about the Rasch model, see references in Appendix G.

In order to determine an appropriate scale transformation, three questions had to be addressed. First, what constant would be used as a multiplication factor? Second, what additive constant would be used to eliminate negative numbers? Third, where should the scale be anchored? Answers to these questions would establish the characteristics of the new score scale.

The multiplication factor was selected to be 20. This would eliminate any decimal values for scores and make it possible to round whole numbers without losing any information. Thus, no two adjacent raw scores would have the same scaled scores.

The additive constant was selected to be 200. This would permit the scale to be expressed entirely with three digits and would not generate a scale identical to any other commonly used scale.

Last, it was determined that the scale would be anchored at the passing score for each subtest; so, even though the subtests vary in difficulty, the person performance level necessary to pass each subtest can be made identical. The passing score value was set at 200.

Setting Passing Score Standards

In preparing for the task of setting the standard for passing the various subtests of the Florida Teacher Certification Examination, an extensive review of the literature was undertaken. Fortunately, many of the most relevant papers have been collected in three publications, and earlier papers are cited in these publications, so that the task of identifying the state of the art was greatly simplified.

The first of these publications is the Florida Journal of Educational Research, Volume 18, 1976. This journal published papers presented in a symposium at the annual meeting of the National Council on Measurement in Education (NCME) in April, 1976. The second publication is the Journal of Educational Measurement, Volume 15, No. 4, Winter 1978, published by NCME, which was a special issue on standard setting. Finally, NCME issued a monograph, Practices and Problems in Competency-Based Measurement, 1979, which was prepared by a task force as the culmination of three years of study and symposia. Many of the papers in this monograph also appear in the other publications.

In developing the procedures to be used in setting passing scores for the FTCE, the materials cited above were very helpful. However, the procedures recommended for the Florida Teacher Certification Examination are not directly adopted from any of the published material. As Bunda and Sanders state in their preface to the standards chapter of the 1979 NCME monograph, "The reader will not find a little jiffy formula in this section which will allow for the setting of standards. The choice of a standard is a complex decision." In his paper in the same monograph, Conaway states,

> "A review of the literature shows that a definitive set of procedures for setting standards in various types of competency-based programs does not exist...it is apparent that practitioners cannot obtain validated standard-setting procedures, either off-the-shelf or out of the literature. Instead, they must devise their own standard-setting procedures, relying upon their own experiences and philosophies in competency-based measurement and upon any guidance they can get from available literature and fellow practitioners."

This is what the Department of Education did for the FTCE. The procedures are described in the following sections of this chapter.

Several procedures for setting standards have been used or proposed in the literature. They generally fall into one of three categories. Each of these categories is presented and discussed in the following pages.

Requirement for Perfection

The "requirements for perfection" method simply selects some passing score less than 100% so that an allowance is made for "careless errors" or "lapses common to all people." Advocates of this procedure often ignore, or are unaware of, the fact that items measuring a given objective or subskill may vary greatly in difficulty because of such irrelevant factors as ambiguity in wording and the closeness of the distractors to the correct (keyed) answer.

The Use of Normative Data

Standards can be set through the use of any of three types of normative data. First, one can simply look at the distribution of actual performance of the group of applicants and set the passing score at some point on that distribution. When only a small percentage is to be selected from a large number of applicants, (as in selecting candidates for medical school, for example) this method works quite well. It is not likely to work so well in selecting only the "incompetent."

A second way of using normative data would be to determine or estimate what proportion of beginning teachers presently employed are "incompetent" on the skills measured by the examination and set the cutting score for the first year in such a way that a specific percentage of applicants is failed. In subsequent years, the cutting score would remain the same, even though fewer and fewer applicants might fail because of improved selection and training in the teacher training institutions.

A third way of using normative data would be to determine the number of new teachers needed each year and set the cut-off score so that a sufficient number of applicants "pass" to provide for the needs.

None of these procedures seemed acceptable as the basis for setting the passing score.

Opinion of Judges

The third method is known as the "jury method" and several different procedures using "judges" have been developed. The judges are usually selected to represent some or all of the important "clienteles" of the test. Most of the procedures described in the literature were developed for use with course examinations and the "jury" was made up of instructors in the courses.

All of these procedures require that the judges conceptualize a minimally or marginally qualified person and keep that person in mind while making their judgements. However, the operational definitions of those terms vary greatly from one situation to another.

The procedure developed by Nedelsky (1954) requires judges to eliminate for each item those options which the "lowest D" student could eliminate. The probability of that student answering the item correctly is the reciprocal of the number of the options left. The sum of the reciprocals is the "guess score" for the "lowest D" student, and the passing score is set somewhere above that point.

Ebel's (1971) procedure requires judges to rate each item as to relevance and difficulty for a "minimally qualified (barely passing) applicant." Then a table is used to assign a "probability" value to each item based on the combination of the ratings on relevance and difficulty. Finally, the average item probability is determined and that becomes the cutting score.

Angoff's (1971) procedure simply requires that each judge state the probability that the "minimally acceptable person" will answer each item correctly. The sum of the probabilities is the cutting score.

All of the preceeding procedures require that judges estimate the difficulty of items. If judges do not have expertise in the area, their estimates may be grossly inaccurate. Furthermore, it seems ridiculous to ask judges to estimate item difficulty when empirical data are available, as they were for the FTCE based on the field test.

In a different kind of procedure, judges are not required to judge item difficulty. Instead, they are provided samples of calibrated items, arranged in order of difficulty, from easy to hard. They are asked to identify the dividing line between items which the minimally qualified examinee should generally answer correctly (probability greater than 50%) and those for which the probability is and should be lower (Draba, 1979).

It was the last procedure that was used for the FTCE. The jury used was the Council on Teacher Education (COTE) which met in Wakulla Springs June 24-26, 1980, and developed recommendations for standards for passing each subtest of the FTCE.

Procedures Used for the Florida Teacher Certification Examination[10]

The judges who recommended the passing score standards for each of the subtests of the FTCE were current and former members of the Council on Teacher Education (COTE). COTE had developed the competency-based model for Teacher Certification, conducted the studies which identified these competencies, and validated subskills and specifications for the examination, so it was very fitting that this group should determine the performance level to be required in order to pass the examination. There were twenty-one judges who assembled in Wakulla Springs on June 24, 1980, to begin the process of setting the passing scores. (See Appendix A for the roster.)

Before the meeting, participants received two documents for study. The first document was the history of the development of the examination. The second document provided technical information to be considered in setting passing scores. The latter document covered the following topics: Overview of methods used in setting standards, issues in setting standards, and recommendations for score scaling.

---

[10]For a complete description of these procedures see COTE Minutes, June 24-26, 1980.

The procedures to be used in setting passing scores were developed in a conference between Dr. Garfield Wilson, Administrator of Teacher Certification, Dr. Dale Lake, Behavioral Science Consultant; and Dr. Annie Ward, Measurement Consultant. Dr. Lake agreed to lead total group discussions and to facilitate group process in order to help the judges arrive at decisions. Dr. Ward provided technical information in both written materials and oral presentations, and Dr. Wilson supervised all administrative details, including invitations to judges and other interested people.

At the beginning of the Wakulla Springs meeting, Dr. Lake discussed the procedures to be used and Dr. Ward discussed the technical information distributed to participants. Staff recommendations for passing scores were given to the participants. It was agreed by the total group that its deliberations would result in recommended passing scores which would balance technical issues, social and political concerns, and practical considerations.

In subsequent sessions, the various subtests were considered one at a time until consensus was reached about the recommendations for each subtest. For each subtest the procedures were as outlined below.[11]

1. In a large group session, Dr. Ward presented the staff recommendation and the rationale for the recommendation. The rationale covered such points as:

   a. General nature of items on the subtest,
   b. Level and range of difficulty of items on the test,
   c. Characteristics of the distribution of scores for the test (i.e., amount and nature of skewness, abrupt changes in slope or breaks in the distribution);
   d. Impact data; and
   e. Research information if available.

2. Judges assembled in mixed role groups. They were given samples of items for the subtest, selected from a full range of difficulty levels. Logit values for each item were provided. Instructions required that the judges sort the items into two groups: (a) those which applicants should be expected to answer correctly more than half, and (b) those which applicants should be expected to answer correctly less than half.[12]

   Then, the judges were to determine the dividing point (in logits) between the two groups. Judges worked independently at first, then attempted to arrive at group consensus. Finally, they compared their decision with the staff recommendation, and either accepted, modified, or rejected that recommendation.

3. As soon as all groups were ready to report, they assembled in a group of the whole. At this session, each group leader presented the recommendation for the group and noted problems or concerns of the group. All these reports were recorded and displayed.

---

[11]See Appendix D for Agenda and Appendix F for materials.

[12]See COTE Minutes, June 24-26, 1980, for a more complete description of the procedures.

4. Judges reassembled into role-alike groups. In these groups they considered political and practical implications of the recommendations. Normative data for various ethnic groups and programs were provided for consideration.

5. After the small group deliberations were completed, the total group was reassembled.

The original plan called for the final recommendation for each subtest to be made at this point. However, with the very first subtest, concern was expressed about setting passing scores on the various subtests in a piecemeal fashion, so it was decided to develop only tentative recommendations at first, then to reconsider all recommendations as a package at the end of the meeting. Therefore, the product of each general meeting was a tentative recommendation and a list of concerns to be re-examined later.

The procedure for the Writing subtest necessarily differed somewhat from that described above for the multiple-choice tests. In the first place, the contractors who had developed the Writing subtest had developed a scoring system which incorporated a standard of "acceptable" or "unacceptable" performance. The judges' task was to examine the application of that standard. Samples of essays were provided to the judges, along with the ratings from three trained raters and, in some cases, a referee. The judges were asked to consider whether they felt the essays had been correctly classified, particularly those which were judged either "Unacceptable" or "Minimally Acceptable."

Recommendations

After all deliberations, the judges recommended acceptance of the staff recommendations of all subtests except the Writing subtest, for which they recommended a slightly higher standard.

The recommended passing scores (expressed in logits) are listed below, with a note about the concerns and considerations which were expressed about each.

Reading: Performance equal to a logit value of 1.40.

Concerns: 1) This standard is above the 75% correct level recommended in research.

2) A high rate of failure for Blacks and some program areas may result.

Consideration: Items are very easy, and there were many perfect scores

Mathematics: Performance equal to logit value of 1.00

Concerns: A high rate of failures, particularly for Black students and some program groups, possibly will result.

Considerations:
1) Items are not too different from high school State Student Assessment Test. Prospective teachers should be expected to do as well as high school students.
2) Mathematics is an easily trained skill, so those who score low can improve their performance in a short period of time.

Professional Education: Performance equal to logit value of .25

Concerns:
1) Many are items possibly irrelevant for the Voc/Tech group.
2) There is a possibility that the scores may be "over-interpreted;" performance assessment in the Professional Education area is important.

Considerations:
1) Items have been judged directly relevant to identified competencies, subskills, and specifications.
2) Items have been developed to achieve a balance across grade levels and subject matter.

Writing: A total score of six or more based on the summed rating of three trained judges using a scale of one to four; at least two of three judges must agree on the acceptability of the writing sample.

Concerns:
1) A high failure rate may occur.
2) It may be difficult for applicants to improve their performance in a short period of time.

Considerations:
1) Even the "minimally acceptable" essays are of poor quality.

These recommendations were submitted to the State Board of Education as a proposed Administrative Rule. The Board adopted the recommendations as Rule on July 30, 1980.

CHAPTER V

BIAS ANALYSES

Test bias has been defined in various ways. One set of definitions relates to the way tests are used and the equity of decisions that are based on the tests. Another set of definitions is related to the content or the format of the items on the test. Bias also is a validity issue because, to the extent that scores on a test reflect group membership rather than the knowledge or skill the test is supposed to measure, the test is invalid. As Shepard, Camilli, and Averill (1980) say, "a test is biased if equally able individuals, from different groups, do not have equal probabilities of success."

Note, however, that differential performance of different groups is not adequate evidence of bias. An example from another field will illustrate this. If a group of undernourished children and a group of well-fed children are weighed, and the well-fed children are found to have higher weights than the undernourished group, it cannot be concluded that the scales are biased. The differential weight (i.e., "performance") reflects accurately the difference in the two groups as to nutrition.

Several methods of investigating bias have been developed and reported in the literature. Many of these procedures are related to decisions made on the basis of the scores. The potential bias of decisions made on the basis of the FTCE will necessarily have to be delayed until criterion data are available (e.g., performance rating on teaching). However, care has been taken to ensure that the content of the test is not biased. This has been accomplished in two steps:

1. In the item development stage, a series of reviews was used to screen items for apparent or potential bias and for offensive materials.

2. Data from the field test were analyzed statistically to identify items which performed differently for different groups. The items flagged by statistical analysis were then inspected for possible biasing elements. The data were analyzed for sex groups, ethnic groups, and program groups.

Developmental Procedures

The procedures used in developing and reviewing test items for potential bias incorporated the practices most frequently recommended in the literature. A set of guidelines for avoiding both the appearance of bias and biasing elements was provided to contractors and incorporated into review forms. (See Appendix E.1.)

Each contractor who wrote items for the FTCE was required to have all items reviewed by at least three experts in the area, and at least one expert was required to represent a minority ethnic group. Review teams were also required to include both men and women. In addition, contractors were required to administer each item on a one-on-one basis to at least three students, one of whom was required to be a minority person. Both sexes and a variety of program areas had to be represented. (See Appendix E.2 for the review form.)

After all items were written they were pilot tested and then reviewed by expert reviewers for potential bias. The bias review team was made up of five members, consisting of two Hispanic and three Black members. Two of the team were male and three were female. The bias review was conducted by Dr. Tom D. Freijo, a measurement expert with considerable experience in item review procedures.

Two Reading passages were flagged as having "sexist" language (i.e., use of generic 'he'); one Mathematics item and one Professional Education item were considered as offensive to one or more ethnic groups. All of these items were replaced.

Statistical Procedures

Statistical methods have been developed to identify items on which groups perform better or worse than their overall performance on the test would predict. The flagged items are then inspected to ascertain whether there are biasing elements in the item. A review of the literature on statistical procedures for detecting potential item bias has been greatly facilitated recently by a paper presented at American Educational Research Association (AERA) by Shepard, Camilli, and Averill (1980). This paper cites several very comprehensive reviews of item-bias detection methods and reports the results of a study in which six of the best known approaches, some with more than one variation, were compared.

The data used in the FTCE studies of potential item bias were collected in the field test in April 1980. The methods for analysis of bias of items on the FTCE were:

1. the transformed Item Difficulty Method (Angoff) introduced in Angoff and Ford (1973);

2. the One-Parameter Item Characteristic Curve Method (Rasch) (Wright, Mead, and Draba, 1976); and

3. correlational studies.

These were the only methods for which sufficient data were available, because although 1,186 people participated in the field test, items were distributed across seven test forms, so N's for some test forms were as small as 150. When these were further sorted into ethnic or program groups, the N's became very minute in some cases. Some groups were combined, when the combination was a logical entity, in order to form groups large enough to analyze.

Before comparisons were made, items with p values lower than .05 and higher than .95 were deleted because for these items group differences are almost certain to be simply chance differences. In addition, items deleted from the pool for technical reasons (poor discrimination or model fit) were not considered. The number of items deleted for these reasons is indicated below:

	No. of items	$p \leq .05$ or $\geq .95$	Technical Deletions	Remaining
Reading	230	141	--	89
Mathematics	102	5	4	93
Professional Education	369	6	29	334

TID Comparisons (Angoff)

The Transformed Item Difficulty (TID) approach relies on an item-by-group interaction definition of bias. Items are flagged for further bias review if they are relatively more difficult for one group than for another. Angoff and Ford (1973) introduced the item-difficulty delta plot method. It is one of the most common methods in practice and is routinely used by the Educational Testing Service as a screening device in test development. Item difficulties or p-values (the proportion of examinees getting the item right) are computed separately for each group, then each p-value is transformed to a z-value corresponding to the $(1 - p)^{th}$ percentile of the standardized normal distribution. The principal axis line is then computed. Items which deviate most from this line (measured as a perpendicular distance) are flagged as potentially biased. The direction of the difference indicates the group for which the item is relatively easier.

The following comparisons were made for all items using the Angoff Method:

Sex Group
Male to Female

Ethnic Group
Black to White
White to Hispanic

Program Groups
Secondary Programs to All Other Programs
Elementary Programs to All Other Programs
Vocational Programs to All Other Programs
Special Education Programs to All Other Programs

An item was flagged for review if the "standardized bias" value was ±2.00 or greater, indicating that the delta difference for the two groups was twice the standard error.

Of the 89 Reading items considered, 18 were flagged for further in-depth review. Performance on four of these favored Secondary Education majors, and performance on five others favored Special Education majors. An inspection of these items does not provide any obvious reason why these items should be relatively easier for these groups. For the other flagged items there was no consistent pattern.

Thirteen of the 93 Mathematics items were flagged for inspection. There was a tendency for the flagged items to favor males (3 items), Secondary Education majors (4 items), and Vocational Education majors (4 items). Inspection of the flagged items did not reveal any biasing elements, except that one which favors Vocational Education majors has a vocational class setting; however, items with elementary school settings did not favor elementary majors. The other flagged items involve simple computation of fractions or decimals. It seems unreasonable to think that the items are really biased toward or against any group.

Twenty-one of the Professional Education items were flagged for review. Five items favored Secondary Education students. These were all related to either Competency 6 (Personal/Social/Academic Development) or Competency 14 (Diagnosis and Assessment). Four items favored Special Education majors. One of these was related to Competency 6, the others were related to Competencies 11 and 12 (Teaching Methodology). Three of the four items which favored Vocational Education majors were from the Teaching Methodology area and the fourth was related to student assessment. Five items favored Hispanics, two favored Whites, and one favored Blacks. There was no simple explanation for these results, since the items cut across all content areas. Furthermore, there was no feature of the context which seemed to be related to any ethnic group. The setting for the items covered five different subject areas, and there was no specification of sex or ethnic variables in any of the items.

In summary, a few of the flagged items favor certain program groups (Secondary, Special, and Vocational Education), but none are biased against any ethnic or sex groups. The greatest percentage of flagged items were Mathematics items and involved very basic mathematical procedures (e.g., addition, subtraction, and division of fractions; percent; ratio; etc.). The Reading and Professional Education items which favored Special Education, Secondary, and Vocational Education students tended to come from the Educational Psychology field. Rather than reflecting bias, it is probable that these results reflect current curricular emphasis for these program areas.

Rasch Analyses

Rasch model methods for identifying biased items are based on the concept that examinees with the same ability should have the same probability of getting an item correct regardless of group membership. After difficulty parameters are equated to the same scale, the most straightforward index of bias is a simple difference in the difficulty parameters estimated separately for two groups.

The Rasch analysis for item bias was limited to linking items, e.g., the subsampled items which were common to all forms of the field test. These common items had been selected to represent the pool of items for each of the subtests. Using the common items made it possible to make comparisons for groups which were too small for intra-form comparison. The number of items involved was as follows:

Reading:	2 passages, 20 items
Mathematics:	15 items
Professional Education:	20 items

The following comparisons were made using the Rasch Model:

Sex Group
Male to Female

Ethnic Group
Black to White
Hispanic to White
Black to Hispanic

Program Groups
Elementary to Secondary
Elementary to Vocational
Elementary to Special Education
Secondary to Special Education
Secondary to Vocational
Special Education to Vocational

An item was flagged as potentially biased if the differences in difficulty logits exceeded ±2 standard deviations. A summary of the results follows.

Reading Subtest Assessing bias for items in the Reading subtest was complicated by the extreme easiness of the test. Of the 20 items which were used as links, only three items (2, 13, and 18) had p values below .95, and one of those (13) had a p value of .90. Therefore, it was decided to limit consideration of bias for only those three items.

Only one of the 30 comparisons (3 items x 10 comparisons) was flagged, less than the number to be expected by chance alone. Inspection of the item and the passage in which it was imbedded did not reveal any obvious biasing element.

Mathematics Subtest None of the 15 linking items for Mathematics was eliminated because the p value was greater than .95, so the group comparisons were made on 15 items, making a total of 150 comparisons. Seven of these differences were statistically significant at the .05 level, exactly the number which would have been expected by chance. Furthermore, the direction of the differences varied across all groups with one each favoring males, white, black, Secondary, Elementary, Special Education, and Vocational.

Professional Education Subtest Of the 20 common items for Professional Education, two were removed from the item pool for technical reasons; none were eliminated because of extreme easiness. Therefore, there were eighteen (18) items involved in the comparison, providing 180 comparisons. Only six (6) of these differences were significant at the .05 level, less than the number expected by chance. These were distributed in favor of Hispanics (2), Special Education, Secondary (2), and Elementary.

In summary, the Rasch analysis did not reveal any items which could be confirmed to be biased.

Correlational Analyses

Correlations were computed for the Rasch item difficulties for each pair of groups for the Mathematics and Professional Education subtests. The Reading subtest was not included in this analysis because of the small number of items involved.

The correlations are reported in Table 5.1. For Mathematics, the correlations range from .710 (Special Education-Vocational Education) to .914 (Elementary-Secondary) with a median of .864. For Professional Education the correlations range from .887 (Elementary-Secondary) to .976 (Male-Female) with a median of .930.

These correlations indicate that the item difficulties for the Professional Education subtest are generally in the same order for all groups. For Mathematics, there is slightly more variation between groups, with both of the lowest correlations involving Special Education majors.

Summary

The Florida Teacher Certification Examination was extensively reviewed for cultural, sexual, and racial or ethnic bias. The bias review procedures included statistical procedures and reviews of the items by professional educators. The final set of approved items are free of bias.

TABLE 5.1

CORRELATIONS OF ITEM DIFFICULTIES
BETWEEN VARIOUS GROUPS

Mathematics	Groups	r
	Male - Female	.849
	Black - White	.904
	Hispanic - White	.877
	Black - Hispanic	.814
	Elementary - Secondary Education	.877
	Elementary - Vocational Education	.840
	Elementary - Special Education	.914
	Secondary - Special Education	.796
	Secondary - Vocational Education	.889
	Special Education - Vocational Education	.710

Professional Education	Groups	r
	Male - Female	.976
	Black - White	.934
	Hispanic - White	.932
	Black - Hispanic	.903
	Elementary - Secondary Education	.887
	Elementary - Vocational Education	.926
	Elementary - Special Education	.933
	Secondary - Special Education	.928
	Secondary - Vocational Education	.944
	Special Education - Vocational Education	.905

APPENDIX A

Professional Contributions (Personnel)

A. Professional Contributions

Many people have participated in the development of the Florida Teacher Certification Examination. Some of these will remain anonymous (e.g., those who participated in the pilot test or field testing). Those with direct responsibilities are listed below:

Director:

Dr. Garfield Wilson, Director of Teacher Education
Department of Education

Measurement Consultants:

Dr. Jacob Beard, Florida State University
Dr. Annie Ward, University of South Florida
Dr. Howard Stoker, Florida State University
Mr. Robert Feinberg, University of Florida

Staff Assistants:

Wayne Harris (1978-79)
Christian Doolin (1979-80)

Competency Survey:

William Davies, Director of Federal Programs and Grants,
Polk County Schools

Preparation of Specifications Bulletin and Technical Manual:

Dr. Annie Ward, Measurement and Research, University of
South Florida

Reviewers of Technical Manual:

Dr. Robert Ebel, Michigan State University
Dr. William Mehrens, Michigan State University
Dr. Roger Wilk, University of South Florida
Dr. Howard Stoker, Florida State University
Dr. Jacob Beard, Florida State University
Dr. Thomas Fisher, Department of Education
Dr. Garfield Wilson, Department of Education
Dr. Kenneth Loewe, Department of Education

COTE Members (1978-80)

Classroom Teachers

William Moore, Seminole County
Quintilla Lynch, Columbia County
Teresa Mention, Dade County
Laurie Murray, Duval County
Anna Jones, Duval County
Ida Weimer, Brevard County
Mary J. Thompson, Sarasota County
Marjorie Head, Broward County
Fannie Williams, Broward County
Walter Johnson, Brevard County
Thomasena Lindsey, Dade County

School Administrators

Ray Pottorf, Superintendent, Lee County
Dr. Donald Gill, Superintendent, Volusia County
James Reese, Principal, Escambia County
John Lizer, Director of Staff Development, Hillsborough County
Lonnie Coleman, Principal, Dade County
Idella Moss, Director of Teacher Education Center, Sarasota County
Jerry Sisk, Principal, Baker County
Samuel Hunter, Assistant Superintendent, Personnel, St. Lucie County
Dr. Cecil Carlton, Jr., Superintendent, Taylor County

Higher Education

Dr. Robert Wiegman, Florida Atlantic University
Dr. Marian Calway, Florida Southern College
Dr. Willis Holcombe, Broward County Community College
Dr. Marcia Mann, University of South Florida
Dr. Eva Wanton, Florida A & M University
Dr. Paul George, University of Florida
Dr. James Richburg, Valencia Community College
Dr. Barney Stoutamire, University of West Florida

Citizens

Beebe White, Volusia County
Geraldine Kalvin, Collier County
John Espey, Pinellas County
Ethel Beckham, Dade County

Ad Hoc Task Force

Dr. Jacob Beard, Professor, Measurement and Research, Florida State University
Dr. William E. Blank, Assistant Professor, Vocational Education, University of South Florida
William Davies, Director, Federal Programs, Polk County
Linda Denmark, Teacher, Polk County
Tom Fackender, Teacher, Leon County
Lewis Jones, Teacher, Leon County
Mary Johnson, Teacher, Polk County
Quintilla Lynch, Teacher, Columbia County
Mary Newton, Teacher, Polk County
Susan Nordeng, Teacher, Bay County
Dr. Daniel Purdom, Professor, Curriculum, University of South Florida
Dr. Annie Ward, Professor, Measurement and Research, University of South Florida

Developers of Item Specifications

Reading: Project Director: Dr. F. J. King, FSU
Project Staff: Dr. John Hills, Carol Lynch-Brown, Dr. Carolyn Schluck, Dr. Lawrence Hafner
Reviewers: Dr. William Castine, FAMU; Dr. Al Lowe, USF

Writing: Project Directors: Dr. Gordon Brossell and Dr. James Hoetker
Project Staff: Dr. Lucy Hayward
Reviewers: Dr. Nancy McGee, UCF; Dr. Dan Kelly, UF

Mathematics: Project Director: Dr. Tom Denmark, FSU
Project Staff: Dr. Janice Flake, Dr. Walter Dick

Reviewers:

Roy Buldoc, UF
Doyle Casteel, UF
Arnold Cheyney, UM
Barry Greenberg, FIU
Sandra K. Kirk, UNF
Ronald LaFaro, BC
Lennie Middleton, UM
H. E. Steiner, USF
Robert Tests, FIU
John Bullock, USF
W. H. Castine, FAMU
Gordon Eade, UWF
Susan Homan, USF
R. E. Lee, UWF
Daryle May, JU
Seaton Smith, UWF
Barbara Stephenson, JU
Nick Vigilante, FIU
Ernie Burgess
Richard Burnette, Fla. So.
Paul Fleming, Fla. So.
Mike Hynes, UCF
Shirley Lee, BCC
Mary Mercer, FAMU
Judith Staver, JU
Ann Taylor, BC
Allen Wuertz, Fla. So.
Arlene Brett, UM

Professional Education

Personal, Social, and Academic Development:

Project Directors: Dr. Gary Peterson and Dr. Steven Rollin, FSU
Project Staff: Dr. Dan Kelly, FSU; Dr. Steven Rollin, FSU; Peter Linkway; Gene McBride; Michael Kormanicki; Kevin Murphy; Robert Conners; Ellen Hume; John Rushbrook

Reviewers: Dr. Daniel Purdom, USF; Dr. Raoul Arreola, FSU; Dr. Herbert Alexander, FAMU; Sara Butzin, Mary Calhoun, Brenda Columbo, James Pitts, (all teachers, FSU Developmental Research School)

Technical Skills:

Project Directors: Dr. Anna Nelson and Dr. Willard Nelson, FAU
Reviewers: Dr. Charles Dziuban, USF; Dr. Annie Ward, USF

Administrative Skills:

Project Director: Dr. Colleen Ryan, FIU
Project Staff: Dr. Barry Greenberg, FIU; Dr. Marisal Gavilon
Reviewers: Dr. Jim Wells, FIU; Dr. Charles Dziuban, UCF; Dr. Tim Sullivan, UCF

Developers of Items

Reading:

Project Director: Dr. F. J. King, FSU
Project Staff: Carol Lynch-Brown
Reviewers: Dr. John Hills, Dr. Laurence Hafner, FSU

Writing:

Project Directors: Dr. Gordon Brossell, Dr. Jim Hoetker, FSU
Project Staff: Barbara Ashe
Reviewers: Dr. Dan Kelly, UF; Dr. Nancy McGee, UCF; Dr. Dwight Burton, FSU (referee); Pamela Laws, Tallahassee Community College; Linda Clarke, Lincoln High School; Carol Gray, Leon High School

Mathematics:

Project Director: Dr. Tom Denmark, FSU
Project Staff: Dr. Robert Kalin, Dr. Janice Flake, Dr. Walter Dick
Reviewers: Dr. William Castine, FAMU; Dr. Andria Troutman, USF; Ms. Marie Boyd, Escambia Co. Teacher; Richard DeAugiro, Dade Co. Teacher

Professional Education

Personal, Social, and Academic Development:

Project Directors: Dr. Gary Peterson, Dr. Dan Kelly, FSU
Reviewers: Dr. William Piotrowski, Leon Co. Schools; Teachers of Leon Schools as follows: Jeff Patterson, Elem. Teacher; Winston Farrell, Principal, Sealy Elem.; Betty Watson, High School Math; George Ortega, High School; Frances Stallworth, Middle School; Carol Senesac, Elem.; Rodney Falmlen, High School; Peter Linkway, High School

Teaching and Learning:

Project Director: Dr. Walter Wager, FSU
Project Assistant: Joan Fox
Reviewers: Dr. Ivy Mitchell, DRS, FSU; Ms. Linda Marts, DRS, FSU; Dr. John Hansen, FSU; Dr. Walter Dick, FSU;

Diagnosis and Assessment:

Project Directors: Dr. Anna Nelson and Dr. Willard Nelson, FAU
Reviewers: Dr. Helen Franke, Palm Beach Co. Schools; Mary C. Greene, Dr. George Mouly, UM; Thais Villanueva, Teacher, Boca Raton Elem. School; Frances Carter, Teacher, Addison Mizner Elem. School; Dr. Murray I. Gellen, FAU; Frank Martinez, Teacher, Boca Raton Middle School; Dr. William V. Meredith

Classroom Management/Administrative Skills:

Project Director: Dr. Ronald Peake, UWF
Project Staff: Dr. Dorothy Diamond, Dr. Helen James, Dr. Hilda Schaer, Mrs. Carlisle Semmes
Reviewers: Dr. Phila Crane, UWF; Dr. Omega Gardner, UWF; Mrs. Hazel Scales, Assist. Principal; Mrs. Jo Tarwater, Elementary Teacher; Dr. James Zafirau, UWF

Item Review:

Director: Dr. Garfield Wilson, DOE
Coordinator: Dr. Annie Ward, USF
Consultants: Dr. William Mehrens, Professor, Educational Measurement, Michigan State University
Dr. Howard Stoker, Professor, Educational Research, Florida State University
Dr. Tom D. Freijo, (Formerly Professor, University of South Florida) President PDE Associates, Tampa
Staff Assistants: Christian Doolin, Clara Combs, William Green, Joan Fox

Reviewers:

Ms. Mary Thompson, Teacher, Venice High School
Ms. Helen Goodson, Teacher, Seminole Co.
Mr. Henry Fraze, Teacher, Northeast High School, Pinellas Co.
Mrs. LeRosa Smith, Assistant Superintendent for Personnel, Volusia Co. Schools
Dr. Bruce Hall, Chairman, Measurement & Research Department, College of Education, University of South Florida
Dr. Dan Purdom, Professor, Curriculum, University of South Florida
Dr. Kenneth Loewe, Student Assessment, Department of Education
Dr. Gabriel Valdes, Consultant, Foreign Language & Bilingual Education, Department of Education
Ms. Lois Plaag, Teacher, Chamberlain High School, Hillsborough Co.
Dr. James Beck, College of Education, Florida A & M University
Dr. Paul Gallagher, Assistant Vice President, Faculty Relations, Florida International University
Dr. Mildred Murray, Coordinator of Federal Programs Evaluation, Hillsborough County Schools
Mrs. Margaret Wolfe, VIS Teacher, Dempsie Brewster Elem. School, Volusia County
Mr. Norman Sank, Teacher, Dade Co. Schools
Dr. Annette Goin, Acting Chairperson, Division of Education, Bethune-Cookman College
Dr. Marian Calway, Director of Teacher Education, Florida Southern College
Ms. Mary Newton, Teacher, Kathleen High School, Polk County
Mr. Larry Wantuck, Math Supervisor, Broward Co. Schools
Dr. Nelson Towle, Coordinator, Media & Instructional Material, Sarasota County Schools
Ms. Quintilla Lynch, Teacher, Columbia Co. Schools
Dr. Linda Crocker, Chairperson, Reasearch & Found. of Education, College of Education, University of Florida
Dr. Carolyn Garwood, Associate Dean for Instruction, College of Education, University of Miami

Field Test:

Coordinator: Dr. Thomas Fisher, Administrator, Student Assessment, Department of Education

Test Sites	Administrators
Bethune-Cookman	Dr. Annie Ward, University of South Florida
Florida Agricultural and Mechanical University Florida State University	Dr. Thaxton Springfield, Leon County Schools
Florida Atlantic University Florida International University Florida Memorial College Palm Beach Atlantic College University of Miami	Dr. Robert Pickens, Palm Beach County Schools

Test Sites (Cont.)

University of Florida	Dr. Mel Lucas, Alachua County Schools
University of Central Florida	Dr. Russell J. Crouse, Orange County Schools
Jacksonville University	Joseph R. Wallace, Duval County Schools
St. Leo College University of South Florida	Dr. Susan Turner, Hillsborough County Schools
Sarasota Vo-Tech Center	Dr. Rick O. Nations, Sarasota County Schools

Standard-Setting Panel

Coordinators: Dr. Garfield Wilson, Department of Education
Dr. Dale Lake, Organizational Development Consultant
Dr. Annie Ward, University of South Florida

Preparation and Presentation of Technical Materials: Dr. Annie Ward
Conduct of Group Discussions: Dr. Dale Lake

COTE Members:

Dr. Marian Calway, Director of Teacher Education, Florida Southern College
Lonnie Coleman, Principal, Dade County
Dr. Donald Gill, Superintendent of Schools, Volusia County
Marjorie Head, Teacher, Broward County
Dr. Willis Holcombe, Provost, Broward County, Community College
Geraldine Kalvin, School Board Member, Collier County
Dr. John Lizer, Director of Staff Development, Hillsborough County
Teresa Mention, Teacher, Dade County
Bill Moore, Teacher, Seminole County
Laurie Murray, Teacher, Duval County
James E. Reese, Principal, Escambia County
Dr. Barney Stoutamire, Director of Teacher Education, University of West Florida
Ida Weimer, Teacher, Brevard County
Fannie Williams, Teacher, Broward County

Former COTE Members:

Hilda Coursey, Program Planning Specialist, Walton County
James Gant, Dean of the College of Education, Florida State University
Quintilla Lynch, Teacher, Columbia County
Irma Norman, Lay Member
Bob Richburg, Vice President Valencia Community College
Mary Thompson, Teacher, Sarasota County
Allan Keeth, School Board Member, Seminole County

Observers:

Chris Doolin, DOE
Fred Milton, DOE
John Staples, DOE
Shelley Boone, DOE
Ralph Turlington, DOE
Claudia Thomas, FTP/NEA
Mrs. Jimmy Reese, Escambia County
Mrs. Barbara Keeth, Seminole County
Mrs. Ralph Turlington, Tallahassee, Fl

APPENDIX B

Evaluation Strategies

for

Florida Generic Competencies and Subskills

Florida Essential Generic Competencies
Evaluation Strategies

KEY: x - Minor Evaluation Emphasis
xx - Major Evaluation Emphasis

	Performance Evaluation			
	Written Exam	Content Area Application	Generic Application	Deleted
1. *Demonstrate the ability to orally communicate information on a given topic in a coherent and logical manner.*			*xx*	
a. Utilizes the principles of simplicity and clarity in organization of oral presentation.			xx	
b. Uses standard English in oral communication.			xx	
c. Uses vocabulary suitable to the topic and audience.		x	xx	
d. Speaks with a volume and pace that promotes comprehension.			xx	
e. Provides verbal and non-verbal cues to the organizational structure of the oral message.			xx	
f. Provides relevant examples that illustrate oral content.			xx	
2. *Demonstrate the ability to write in a logical, easily understood style with appropriate grammar and sentence structure.*	*xx*		*x*	
a. Differentiates between formal and informal written English and demonstrates ability to use both forms.				x

KEY: x - Minor Evaluation Emphasis
xx - Major Evaluation Emphasis

	Performance Evaluation			
	Written Exam	Content Area Application	Generic Application	Deleted
b. Uses language at the level appropriate to the topic and reader.	x	xx	x	
c. Comprehends and applies basic mechanics of writing; spelling, capitalization and punctuation.	xx			
d. Comprehends and applies appropriate sentence structure.	xx			
e. Comprehends and applies basic techniques for the organization of written material.	xx			
f. Comprehends and applies standard English usage in written communications.	xx			
3. *Demonstrate the ability to comprehend and interpret a message after listening.*			xx	
a. Accurately follows multi-step oral directions.			xx	
b. Listens effectively for the organization, main idea, subordinate ideas and details of a message.		x	xx	
c. Listens effectively in order to identify relevant information and propaganda techniques.		x	xx	
d. Listens effectively in order to draw inferences.		x	xx	

KEY: x - Minor Evaluation Emphasis
xx - Major Evaluation Emphasis

	Written Exam	Performance Evaluation: Content Area Application	Performance Evaluation: Generic Application	Deleted
e. Summarizes the message after listening.		x	xx	
f. Comprehends both standard and non-standard English language used by students.			xx	
4. *Demonstrate the ability to read, comprehend, and interpret orally and in writing, professional material.*	*xx*		*x*	
a. Identifies and evaluates relevant professional material.	xx	x	x	
b. Understands basic statistical terminology (such as: mean, median, mode).				x
c. Demonstrates literal reading skills (such as recognizing main idea, details, sequencing, comparison, and contrast).	xx		x	
d. Demonstrates interpretive reading skills (such as: predicting outcome, drawing conclusions, making generalizations).	xx	x	x	
e. Demonstrates critical reading skills (such as: recognition of relevant and irrelevant information, propaganda techniques, and fallacies in reasoning).	xx	x	x	

KEY: x - Minor Evaluation Emphasis
xx - Major Evaluation Emphasis

	Performance Evaluation			
	Written Exam	Content Area Application	Generic Application	Deleted
f. Produces a logical summary interpretation of the results of research in professional material.				x
5. *Demonstrate the ability to comprehend and work with fundamental mathematical concepts.*	*xx*			
a. Adds, subtracts, multiplies, and divides whole numbers, decimals, and fractions.	xx			
b. Demonstrates the meaning and use of fractions and percents.	xx			
c. Represents and interprets data using charts, tables, graphs, and maps.	xx			
d. Solves measurement problems involving length, area, volume, capacity, weight, time, and temperature, using U. S. customary and metric units.	xx*			
e. Applies mathematical skills to solve real world problems.	xx	x		
f. Identifies geometric forms and relationships.				x

*Not validated in professional survey but was retained because of statutory requirement (229.841 F.S.).

KEY: x - Minor Evaluation Emphasis
xx - Major Evaluation Emphasis

		Performance Evaluation			
		Written Exam	Content Area Application	Generic Application	Deleted
6/19.	*The ability to comprehend patterns of physical, social, and academic development in students and to counsel students concerning their needs in these areas.*	*x*		*xx*	
	a. Demonstrates knowledge of basic principles of human growth and development.	xx		x	
	b. Obtains knowledge of students through classroom tests, teacher observations, and student records to contribute to understanding of student needs.	x	x	xx	
	c. Explains test data and other classroom evaluations to students in relation to their educational needs.				x
	d. Applies motivational techniques to encourage students to be achievement-oriented and goal-directed.	x	x	xx	
	e. Assists the student in relating achievements and interests to aptitude and ability.		x	xx	

KEY: x - Minor Evaluation Emphasis
xx - Major Evaluation Emphasis

	Written Exam	Performance Evaluation: Content Area Application	Performance Evaluation: Generic Application	Deleted
f. Demonstrates knowledge of alternative school and community resources for students who have special needs.				x
g. Assists students in developing individual learning activities.		x	xx	
h. Develops student awareness of career opportunities using school and community resources.				x
7. *Diagnose the entry level knowledge and/or skills of students for a given set of instructional objectives using diagnostic tests, teacher observation, and student records.*	*xx*	*x*	*xx*	
a. Selects the specified knowledge or skill to be diagnosed, and determines the most appropriate method for conducting the diagnosis.	xx	x	x	
b. Selects or constructs a test to diagnose student learning needs.	xx	x	x	
c. Uses classroom observation techniques to diagnose student learning needs.	x	x	xx	
d. Uses information from student records to diagnose student learning needs.	x		xx	

KEY: x - Minor Evaluation Emphasis
xx - Major Evaluation Emphasis

	Performance Evaluation			
	Written Exam	Content Area Application	Generic Application	Deleted
e. Interprets results obtained from diagnostic tests, teacher observation, and information from student records.	xx	x	x	
8. *Identify long-range goals for a given subject area.*		*x*	*xx*	
a. Identifies state and district long-range goals.				x
b. Formulates subject area goals consistent with state and district goals and student needs.	Addressed in (9a)			
9. *Construct and sequence related short-range objectives for a given subject area.*	*xx*	*xx*	*x*	
a. Identifies knowledge, skills, and attitudes to be attained for a subject area.	xx	xx	x	
b. Constructs or adapts short-range objectives for identified knowledge, skills, and attitudes.	xx	xx	x	
c. Sequences short-range objectives consistent with commonly accepted principles of learning.	xx	xx	x	
10. *Select, adapt, and/or develop instructional materials for a given set of instructional objectives and student learning needs.*	*x*	*x*	*xx*	

KEY: x - Minor Evaluation Emphasis
xx - Major Evaluation Emphasis

	Written Exam	Performance Evaluation: Content Area Application	Performance Evaluation: Generic Application	Deleted
a. Determines desirable characteristics of materials based on objectives and student learning needs.	xx	x	xx	
b. Locates and evaluates available instructional material.	x	x	xx	
c. Selects materials to assist students in mastering an objective.	x	x	xx	
d. Demonstrates techniques for modifying materials to assist students in mastering an objective.	x	x	xx	
e. Determines materials to be developed based upon existing resources and student needs.	x	x	xx	
f. Identifies and selects resources needed for materials development.	x	x	xx	
g. Designs and constructs materials based on instructional objectives, student needs, and available resources.			xx	
h. After use, evaluates the effectiveness of instructional materials in accomplishing objectives, and revises accordingly.	x	x	xx	
11. *Select/develop and sequence related learning activities appropriate for a given set of instructional objectives and student learning needs.*	*x*	*xx*	*x*	

KEY: x - Minor Evaluation Emphasis
xx - Major Evaluation Emphasis

	Performance Evaluation			
	Written Exam	Content Area Application	Generic Application	Deleted
a. Comprehends basic principles of human growth and development.			x	
b. Identifies conditions that affect learning.	xx	x		
c. Identifies alternative activities to achieve an objective.	xx	xx	x	
d. Selects an appropriate learning activity to achieve an objective.	xx	xx	x	
e. Combines appropriate learning activities into an instructional sequence.	xx	xx	x	
12. Establish rapport with students in the classroom by using verbal and/or visual motivational devices.	x	x	xx	
a. Secures the attention of students through appropriate devices.	x		xx	
b. Relates instructional objectives and activities to interests, capabilities, and experiences of students.	x	x	xx	
c. Informs students about objectives, subsequent learning tasks, and performance expectations.	x	x	xx	
d. Explains choices and limitations of possible learning activities.		x	xx	

KEY: x - Minor Evaluation Emphasis
xx - Major Evaluation Emphasis

	Performance Evaluation			
	Written Exam	Content Area Application	Generic Application	Deleted
e. Alters instructional strategies during learning activities based on student responses and other factors.	x	x	xx	
f. Relates students' and teacher's experiences, thoughts, and feelings to learning activities.		x	xx	
g. Uses reinforcement techniques to assist in student motivation.	x		xx	
h. Uses media to secure interest and maintain attention.	x	x	xx	
i. Uses student products and talent to secure interest and maintain attention.		x	xx	
13. *Present directions for carrying out an instructional activity.*	x	x	xx	
a. Selects appropriate means for presenting directions.	xx	x	xx	
b. Secures attention of students for the purpose of giving directions.		x	xx	
c. Informs students of objectives, assessments, and performance standards.		x	xx	
d. Informs students of the sequence and nature of learning activities to achieve the objectives.		x	xx	

KEY: x - Minor Evaluation Emphasis
xx - Major Evaluation Emphasis

	Written Exam	Performance Evaluation: Content Area Application	Performance Evaluation: Generic Application	Deleted
e. Identifies materials for a learning activity and explains their use.	x	x	x	
f. Determines if students understand directions.		x	xx	
g. Clarifies directions by responding to student questions.		x	xx	
14. *Construct or assemble a classroom test to measure student performance according to criteria based upon objectives.*	xx	x	x	
a. Identifies uses of basic types of classroom tests and assesment techniques.	xx	x	x	
b. Identifies appropriate uses of norm-referenced and criterion-referenced testing.	xx*	x	x	
c. Given an objective, specifies knowledge and skills to be assessed.	xx	x	x	
d. Selects appropriate assessment techniques to evaluate mastery of an objective.	xx	x		
e. Determines limitations, constraints, and requirements for administering tests.	xx	x		

*Not validated in professional survey but was retained because of statutory requirement (229.575 F.S.).

KEY: x - Minor Evaluation Emphasis
xx - Major Evaluation Emphasis

	Written Exam	Performance Evaluation: Content Area Application	Performance Evaluation: Generic Application	Deleted
f. Constructs and identifies test items and tasks that evaluate mastery of an objective.	xx	xx	x	
g. Identifies criteria for standards of performance.	xx	xx	x	
h. Assembles test components including test items, directions, and scoring keys.		x	xx	
i. Evaluates and/or revises tests on the basis of validity, reliability, and student responses.	x	x	x	
15. *Establish a set of classroom routines and procedures for utilization and care of materials.*	x	x	xx	
a. Involves students in developing classroom routines and procedures for utilization and care of materials.		x	xx	
b. Determines the type and amount of materials necessary to complete classroom assignments.	x	xx	x	
c. Organizes an effective system for placement and distribution of materials in the classroom.	x	x	xx	

KEY: x - Minor Evaluation Emphasis
xx - Major Evaluation Emphasis

	Performance Evaluation			
	Written Exam	Content Area Application	Generic Application	Deleted
d. Organizes and arranges a center that will serve as a focus of interest for student learning (such as a bulletin board, display table, or exhibit).		x	xx	
e. Identifies physical elements and arrangements in the classroom that directly affect learning.	xx	x	x	
f. Involves students in developing routines and procedures for physical movement in the classroom.	x	x	xx	
g. Arranges classroom furniture and equipment to accomodate selected teaching strategies.	x	x	xx	
h. Identifies approved procedures for movement of students in emergencies that can be anticipated.	xx	x		
16. *Formulate a standard for student behavior in the classroom.*	x	x	x	
a. Identifies approved safety procedures and incorporates them into a standard for student behavior in the classroom.	xx	x	x	

KEY: x - Minor Evaluation Emphasis
xx - Major Evaluation Emphasis

	Written Exam	Performance Evaluation: Content Area Application	Performance Evaluation: Generic Application	Deleted
b. Identifies and incorporates socially accepted norms (such as mutual respect, consideration of others, courtesy) into a standard for student behavior in the classroom.	xx		xx	
c. Identifies characteristics of the student population (such as age and maturity) that need to be considered in formulating a standard for student behavior in the classroom.	xx	x	x	
d. Establishes a realistic standard of behavior that has potential for consistent application.	x	x	xx	
e. Identifies and incorporates state and local policies into a standard for student behavior in the classroom.	x		xx	
17. *Identify causes of classroom misbehavior and employ a technique(s) for correcting it.*	*x*	*x*	*xx*	
a. Identifies factors of the physical environment that affect student behavior.	xx	x	x	
b. Identifies social and emotional characteristics of the teacher that affect student behavior.	xx	x	x	

KEY: x - Minor Evaluation Emphasis
xx - Major Evaluation Emphasis

	Written Exam	Performance Evaluation: Content Area Application	Performance Evaluation: Generic Application	Deleted
c. Identifies physical, social, and emotional characteristics of the student that affect student behavior.	xx	x	x	
d. Identifies out-of-school factors that affect student behavior.	xx	x	x	
e. Identifies aspects of instructional procedures and techniques which affect student behavior.	xx	x	x	
f. Demonstrates effective techniques and strategies for managing student behavior.	xx	x	x	
g. Uses selected verbal and non-verbal techniques for reinforcing and modifying student behavior.		x	xx	
h. Identifies and uses school and community resources for assistance in modifying student behavior.		x	xx	
i. Obtains and utilizes parental assistance for modifying student behavior.		x	xx	
18. *Identify and/or develop a system for keeping records of class and individual student progress.*	*xx*	*x*	*x*	

KEY: x - Minor Evaluation Emphasis
xx - Major Evaluation Emphasis

	Written Exam	Performance Evaluation: Content Area Application	Performance Evaluation: Generic Application	Deleted
a. Constructs a system for recording individual student knowledge and skills progress in a subject area.	xx	x	x	
b. Identifies methods for reporting individual student progress in knowledge and skills in a subject area.	xx	x	x	
c. Identifies methods for recording class progress in knowledge and skills in a subject area.	xx	x	x	
d. Identifies methods for reporting class progress in knowledge and skills in a subject area.	xx	x	x	
e. Demonstrates knowledge of the laws and policies governing the content and use of student records.	xx		x	
19. *(See Competency #6)*				
20. *Identify and/or demonstrate behaviors which reflect a feeling for the dignity and worth of other people including those from other ethnic, cultural, linguistic, and economic groups.*	*x*		*xx*	
a. Creates a learning environment in which students express themselves openly and honestly.		x	xx	

KEY: x - Minor Evaluation Emphasis
xx - Major Evaluation Emphasis

	Performance Evaluation			
	Written Exam	Content Area Application	Generic Application	Deleted
b. Assists students in understanding that individual differences enable each person to make unique contributions to the group effort.	x	x	xx	
c. Demonstrates awareness of cultural differences in dress, beliefs, and practices.	x		xx	
d. Establishes an environment for positive communication and interaction between students from different sociocultural backgrounds.	x		xx	
21. *Demonstrate instructional and social skills which assist students in developing a positive self-concept.*	x	x	xx	
a. Exhibits behavior in the classroom that is empathetic, positive, and reinforcing.		x	xx	
b. Assists students in initiating self-directed learning.	x	x	xx	
c. Assists students in understanding their needs, motives, experiences, and individual value and dignity.	x	x	xx	
d. Selects and uses curriculum materials in accordance with the abilities and mastery levels of individual students.	xx	x	xx	

KEY: x - Minor Evaluation Emphasis
xx - Major Evaluation Emphasis

	Performance Evaluation			
	Written Exam	Content Area Application	Generic Application	Deleted
22. *Demonstrate instructional and social skills which assist students in interacting constructively with their peers.*	*x*	*x*	*xx*	
a. Establishes an environment that permits students to cooperate and share ideas and materials.		x	xx	
b. Assists students in applying constructive criticism in response to each others' work.				x
c. Establishes a learning environment designed to assist students in exhibiting positive interpersonal traits (such as mutual respect and cooperation).		x	xx	
d. Uses techniques that assist students in examining their values, attitudes, and beliefs.	x	x	xx	
23. *Demonstrate teaching skills which assist students in developing their own values, attitudes, and beliefs.*	*x*	*x*	*xx*	
a. Assists students in understanding the need to explore alternative solutions to problems.		x	xx	
b. Establishes teaching strategies that allow students to make choices based on clearly defined consequences.	x	x	xx	

APPENDIX C

Coverage of the Examination

Coverage of the Examination

During the process of review and development described in Chapter I, some subskills were deleted and a few of the competencies and subskills were judged not to be assessable by a written test. Nineteen of the competencies, covering 80 subskills, have been judged to be measurable to some extent by a written test. Appendix B indicates those subskills which were deleted and, for those which remain, it indicates the primary assessment technique to be used.

The Teacher Certification Examination covers those competencies and subskills checked in the first column of Appendix B. The examination consists of four subtests: three in General Education (reading, writing, and mathematics) and one in Professional Education. The competencies to be covered by each subtest are as follows:

Reading	Competency 4
Writing	Competency 2
Mathematics	Competency 5
Professional Education	Competencies 6, 7, 9, 10, 11, 12, 13, 14, 16, 17, 18, 20, 21, 22, and 23.

Examination Schedule

The examination will be scheduled to cover a full day. A typical examination schedule might be as follows:

8:00 - 8:30	Check in, including retakes for Writing Subtest
8:30 - 9:00	General Instructions
9:00 - 9:45	Writing Subtest (Production Task)
9:45 - 10:00	Break (Check in retakes for Mathematics and Reading Subtests)
10:00 - 11:00	Mathematic Subtest
11:00 - 12:00	Reading Subtest
12:00 - 1:30	Lunch
1:15 - 1:30	Check retakes for Professional Education
1:30 - 4:00	Professional Education

Examination Dates

November 22, 1980
April 4, 1981
July 11, 1981

Examination Locations

Eight sites have been identified for the 1980-81 year.

University of North Florida, Jacksonville
University of Florida, Gainesville
Florida State University, Tallahassee

University of West Florida, Pensacola
University of South Florida, Tampa
University of Central Florida, Orlando
Miami Dade Community College, South, Miami
Edison Community College, Ft. Myers

The Subtests

The Florida Teacher Certification Examination is divided into four subtests: Reading, Writing, Mathematics, and Professional Education.

1. Reading:

   a. The Reading subtest uses the multiple-choice CLOZE procedure.

   b. The test consists of eight passages of approximately 100 words, selected from the following sources: (1) textbooks commonly used in required professional education courses, (2) journals and newsletters published by non-specialized teacher organizations, (3) documents for teachers-in-general produced by the State Department of Education, and (4) teachers' manuals for tests in common usage in Florida.

   c. Each passage has ten words deleted. Deleted words are nouns, verbs, adjectives or adverbs, not structural words.

   d. Examinees are asked to select the word to fill each deletion from among four choices which are syntactically equivalent but different in meaning.

2. Writing:

   a. Each examinee writes on one topic, to be selected from two options.

   b. All topics are required to meet these criteria:

      1. Self-explanatory (i.e., clearly and explicitly phrased)
      2. Defined and limited
      3. Familiar to every examinee
      4. Stimulating
      5. Fresh
      6. Of middle emotional ground (i.e., neither too pedestrian nor too sensational)
      7. Non-biased and non-biasing

   c. Scoring is holistic (general impressionistic).

   d. Judges have been specially selected and trained. They are required to reach a specified criterion of agreement with other judges (80%).

   e. Ratings are on a 4-point scale, with the score being the sum of three ratings.

3. Mathematics:

a. The Mathematics subtest consists of approximately 40 multiple-choice items, divided approximately as follows:

Subskills	50% Computation/ Understanding	50% Real World/ Problem Solving	No.	%
(1) Adds, subtracts, multiplies and divides whole numbers, decimals, and fractions.	15	4	19	48%
(2) Demonstrates the meaning and use of fractions and per cents.	3	3	6	14%
(3) Represents and interprets data using charts, tables, graphs, and maps	1	2	3	8%
(4) Solves measurement problems involving length, area, volume, capacity, weight, time and temperature, using U.S. customary and metric units.	1	5	6	14%
(5) Applies mathematical skills to solve real world problems.		6	6	16%
	20	20	40	100%

b. Additional Recommendations:

REAL WORLD

50% items are teacher related and 50% items consumer related/problem solving
33% items can be answered by making a "ball-park" estimate
16% items contain extraneous information
30% items have information presented in a graph, table or drawing
50% items involve two or more steps
16% items are analysis or synthesis tasks

MEASUREMENT ITEMS

50% English and 50% metric with at least 12 items distributed across the following topics: length, weight, capacity, area/perimeter/volume, elapsed time

GENERAL

36% items have "none of the above" as an option as indicated in the item specifications
12% items have "none of the above" as the correct answer (6 Real World, 6 Computation/fact/understanding)

4. Professional Education:

Examinees have 2 and 1/2 hours to complete the Professional Education portion of the exam.

The Professional Education subtest consists of approximately 100 multiple-choice items. The distribution of items will be based on these considerations:

1. All competencies 6-23 are covered. Competency 8 is addressed by the examination items related to Competency Subskill 9a. Competencies 6 and 19 were combined by statute.

2. The extent of the coverage is based on the priorities set by the Professional Task Force.

3. The coverage is based on a balance between the major categories of the content base and the competencies as indicated below:

	CONTENT BASE						
Competencies	Classroom Management	Development of Students	Eval. Recording Reporting Stud. Prog.	Instructional Materials	Instructional Objectives	Learning and Teaching	Totals
6. The ability to comprehend patterns of physical, social, and academic development in students and to counsel students concerning their needs to these areas.	1	10	1			3	15
7. Diagnose the entry level knowledge and/or skills of students for a given set of instructional objectives using diagnostic tests, teacher observation, and student records.			6				6
9. Construct and sequence related short-range objectives for a given subject area.					5		5
10. Select, adapt, and/or develop instructional materials for a given set of instructional objectives and student learning needs.				9			9
11. Select/develop and sequence related learning activities appropriate for a given set of instructional objectives and student learning needs.		7				3	10
12. Establish rapport with students in the classroom by using verbal and/or visual motivational devices.						6	6
13. Present directions for carrying out an instructional activity.						2	2
14. Construct or assemble a classroom test to measure student performance according to criteria based upon objectives.			7		1		8
15. Establish a set of classroom routines and procedures for utilization and care of materials.	7			1			8
16. Formulate a standard for students behavior in the classroom.	6	1					7
17. Identify causes of classroom misbehavior and employ a technique(s) for correcting it.	7	1				1	9
18. Identify and/or develop a system for keeping records of class and individual student progress.			5				5
20. Identify and/or demonstrate behavior which reflects a feeling for the dignity and worth of other people including those from other ethnic, cultural, linguistic, and economic groups.		3					3
21. Demonstrate instructional and social skills which assist students in developing a positive self-concept.						5	5
22. Demonstrate instructional and social skills which assist students in interacting constructively with their peers.						1	1
23. Demonstrate teaching skills which assist students in developing their own values, attitudes, and beliefs.						1	1
TOTALS	21	22	19	10	6	22	100

APPENDIX D

Agendas

D.1 Expert Review Panel

D.2 Standard-Setting Panel

Sunday Evening, January 27 7:00 - 8:30 p.m.

Highlands Room, Tallahassee Hilton, Second Floor

WELCOME - Commissioner Turlington

INTRODUCTION TO THE CERTIFICATION EXAMINATION: Garfield Wilson

General Orientation: Overview, Work Places, Work Plan, Materials

Legal Issues: Judy Brechner, DOE Attorney

Orientation to TASK I: Keying, Traceback, Appropriateness Rating

Monday, January 28

House Office Building, Lower Level

Morris Hall

8:30 a.m. - 11:00 a.m. Task I

11:00 a.m. Turn in materials to Staff Assistants

11:00 a.m. - 12:00 p.m. Orientation Sessions for Task II, Content, Measurement, Bias Reviews

A. White Task Force / Stoker - Room 16
B. Blue, Yellow, Green, Pink Task Forces / Mehrens - Morris Hall
C. Gold Task Force / Freijo - Room 24

12:00 p.m. - 1:00 p.m. LUNCH

1:00 p.m. - 5:00 p.m. Task II

(Materials are arranged in groups. Turn in materials for each section as soon as materials are completed.) Staff will collate materials for each item review and identify items to be rewritten.

Tuesday, January 29

House Office Building

8:30 a.m. - 12:00 p.m. Continue Task II as assigned

12:00 p.m. - 1:00 p.m. LUNCH

1:00 p.m. - 5:00 p.m. Continue Task II, Begin Task III

Staff will begin review of items at noon, Tuesday (or as soon as reviews are completed for a test area).

Wednesday, January 30

8:30 a.m. - 3:00 p.m. Task III / Rewrite, Re-review

77

REVIEW TASK FORCE
HILTON HOTEL

STAFF: Dr. Garfield Wilson, Dr. Annie Ward, Chris Doolin, John Green, Clara Combs, Joan Fox

CONSULTANTS:

Dr. William Mehrens, Professor
Educational Measurement
460 Erichson Hall
Michigan State University
East Lansing, Michigan 48823

Dr. Howard Stoker, Professor
Florida State University (on leave)
Current Assignment:

ETS Atlanta
3445 Peachtree Road, N.E.
Atlanta, Georgia 30326

Dr. Tom Freijo, President
Planning, Development, & Evaluation Associates, Inc.
P.O. Box 17288
Tampa, Florida 33682

REVIEWERS:

Ms. Mary Thompson
Venice High School
Venice, Florida 33595

Ms. Helen Goodson
P.O. Box 1522
Longwood, Florida 32750

Mr. Henry Fraze
6921 17th Lane, North
St. Petersburg, Fla. 33702

Mrs. LeRosa Smith
Asst. Superintendent
Personnel
Volusia County Schools
DeLand, Florida 32720

Dr. Bruce Hall, Chairman
Measurement & Research Dev.
College of Education
Univ. of South Florida
Tampa, Florida 33620

Dr. Dan Purdom, Professor
College of Education
Univ. of South Florida
Tampa, Florida 33620

Dr. Kenneth Loewe
Student Assessment
Department of Education
Tallahassee, Fla. 32301

Dr. Gabriel Valdes
Consultant
Foreign Language and Bilingual Education
Department of Education
Knott Building
Tallahassee, FL 32301

Ms. Lois Plaag
4505 N. Rome Avenue
Tampa, Florida 33603

Dr. James Beck
College of Education
FAMU
Tallahassee, Fla. 32307

Dr. Paul Gallagher
Asst. Vice President
Faculty Relations
Fla. International Univ.
Miami, Florida 33199

Dr. Mildred Murray
Fed. Programs Evaluation
Hillsboro County Schools
707 E. Columbus Drive
Tampa, Florida 33602

Mrs. Margaret Wolfe
VIS Teacher
Dempsie Brewster Elementary School
DeLand, Florida 32720

Mr. Norman Sank
8921 Palm Tree Lane
Pembrook Pines, Fla. 33024

Dr. Annette Goin, Dir.
Teacher Education
Bethune-Cookman College
640 Second Avenue
Daytona Beach, Fla. 32015

Dr. Marian Calway
Florida Southern College
Lakeland, Florida 33803

Ms. Mary Newton
Kathleen High School
Kathleen, Florida 33849

Mr. Larry Wantuck
Math. Supervisor
Broward County Schools
P.O. Box 5408
Ft. Lauderdale, Fla. 33310

Dr. Nelson Towle
Coordinator
Media & Instructional Mat.
Sarasota County Schools
2418 Hatton Street
Sarasota, Florida 33577

Ms. Quintilla Lynch
P.O. Box 336
Branford, Florida 32008

Dr Linda Crocker
Chairperson
Research & Found. of Ed.
College of Education
University of Florida
Gainesville, Fla. 32611

Dr. Carolyn Garwood
Assoc. Dean of Instruction
College of Education
University of Miami
Miami, Florida 33124

78

CONTRACTORS:

Mathematics

Dr. Tom Denmark
Curriculum & Instruction
219 Education Building
Florida State University
Tallahassee, Fla. 32306

Reading

Dr. F. J. King
Research, Development
and Foundations
307-C Stone Building
Florida State University
Tallahassee, Fla. 32306

Writing

Dr. Gordon Brossell
Curriculum & Instruction
213 Education Building
Florida State University
Tallahassee, Fla. 32306

Data Analysis

Dr. Jacob Beard
307 Stone Building
Florida State University
Tallahassee, Florida 32306

Professional Education - Competencies 6, 20, 21, 22, 23

Dr. Gary Peterson
Center for Educational Technology
1-A Tully Gym
Florida State University
Tallahassee, Fla. 32306

Professional Education - Competencies 7, 12, 14

Dr. Willard Nelson
Foundations Department
College of Education
Florida Atlantic Univ.
Boca Raton, Fla. 33431

Dr. Anna Nelson
Foundations Department
College of Education
Florida Atlantic Univ.
Boca Raton, Fla. 33431

Professional Education - Competencies 15, 16, 17, 18

Dr. Ronald Peake
Elem. & Sec. Education
Univ. of West Florida
Pensacola, Fla. 32504

Professional Education - Competencies 9, 10, 11, 13

Dr. Walter Wager
307 Stone Building
Florida State University
Tallahassee, Fla. 32306

THE FLORIDA TEACHER CERTIFICATION EXAMINATION

ITEM REVIEW TASK FORCE

JANUARY 27-30, 1980

State of Florida
Department of Education
Tallahassee, Florida
Ralph D. Turlington, Commissioner
Affirmative action/equal opportunity employer

TASK I GROUPS	NUMBER ITEMS IN PRIMARY GROUP	NUMBER PRIMARY ITEMS PER PERSON	NUMBER OTHER ITEMS (I-M-2 or I-P-2)				READING PASSAGE	TOTAL ITEMS PER PERSON		
			M	A	B	C		1	2	Rdg
			26	26	21	23				
M Mathematics (Yellow) Wantuck, Fraze Sank, Goodson W. Nelson, Murray	102	51	--	9	22	16	3	51	47	3
A Assessment (Green) Hall, Loewe Gallagher, Crocker King, Peterson	115	58	11	--	21	16	3	58	48	3
B Teaching, Learning, Methods, Classroom Procedures (Blue) Purdom, Denmark Calway, Goins Thompson, Lynch Towle, A. Nelson Newton, Valdes	150	41	12	17	--	17	3	41	46	3
C Personal, Social, Academic Development (Pink) Garwood, Beck Fox, Wager Smith, Plaag Peake, Wolfe	124	41	12	13	22	--	3	41	47	3

Task II Groups

GROUPS	ITEMS TO BE REVIEWED 3 Reviews Per Item	ITEM FORMS ITEM FORMS	REVIEW FORMS REVIEW FORMS
asurement- I te) - STOKER II Hall Crocker Murray Gallagher	ALL	Reading Booklets/4 Forms Math Booklets/3 Forms PSA Booklets/3 Forms T & L Booklets/2 Forms AS Booklets/3 Forms D A Booklets/2 Forms Writing Study Report	II-M-R-1 II-M-R-2 II-M-M-1 II-M-M-2 II-M-P-1 II-M-P-2 " " " " " " II-M-W-1 II-M-W-2
Bias (Gold) - FREIJO Valdes Newton Smith Wolfe Beck	ALL	Test Booklets: Reading 4 Math 3 PSA 3 T & L 2 AS 3 D A 2 Writing Instructions	II-B-R II-B-M II-B-P " " " II-B-W
Content - MEHRENS	AS ASSIGNED TO GROUP	Four Groups of Items: Math 102 Group A 115 Group B 150 Group C 124	II-C II-C-M II-C-P II-C-P II-C-P

Group M MATH (Yellow)	Group C PSAD (Pink)	Group B TLM (Blue)	Group A DAT (Green)
Wantuck	Garwood	Purdom	Loewe
Fraze	Plaag	Towle	King
Goodson	Peake	Goins	Peterson
W. Nelson	Wager	Denmark	
Sank	Calway	A. Nelson	
		Thompson	
		Lynch	

AGENDA
COUNCIL ON TEACHER EDUCATION

June 24-26, 1980
Wakulla Springs Inn

TIME	ITEM	WHO	ACTION REQUIRED OR TAKEN
Tuesday, June 24, 1980			
7:00 P.M.	Registration	Chris Doolin	Security Statements
7:15 P.M.	General Session		
	Call to Order	Marian Calway	
	Introductions	Garfield Wilson	
	Introduction to the Process	Dale Lake	
	Presentation of Tentative Recommendation	Annie Ward	
	Identification of Parameters	Dale Lake	
Wednesday, June 25, 1980			
8:30 A.M.	General Session	Annie Ward	Receive technical materials for Recommendation #1
9:00 A.M.	Mixed Role Groups	Group Leaders	Modify Recommendation #1 and Rationale
10:00 A.M.	Break		
10:30 A.M.	General Session	Dale Lake	Develop Second Draft of Recommendation #1
11:15 A.M.	Role - Alike Groups	Group Leaders	Check Recommendation #1 Against Decision Matrix
12:00 Noon	Lunch		
1:00 P.M.	General Session	~~Marian Calway~~	~~Action on Recommendation #1~~
		Annie Ward	Receive technical materials for Recommendation #2

AGENDA
COUNCIL ON TEACHER EDUCATION

June 24-26, 1980
Wakulla Springs Inn

TIME	ITEM	WHO	ACTION REQUIRED OR TAKEN
Wednesday, June 25, 1980 (continued)			
2:00 P.M.	Mixed Role Groups	Group Leaders	Modify Recommendation #2 and Rationale
3:00 P.M.	Break		
3:30 P.M.	General Session	Dale Lake	Develop Second Draft of Recommendation #2
4:15 P.M.	Role - Alike Groups	Group Leaders	Check Recommendation #2 Against Decision Matrix
5:00 P.M.	General Session	~~Marian Galway~~	~~Action on Recommendation #2~~
5:30 P.M.	Dinner Break		
7:00 P.M.	General Session	Annie Ward	Receive technical materials for Recommendation #3
7:30 P.M.	Mixed Role Groups	Group Leaders	Modify Recommendation #3 and Rationale
Thursday, June 26, 1980			
8:00 A.M.	General Session	Dale Lake	Develop Second Draft of Recommendation #3
9:00 A.M.	Role - Alike Groups	Group Leaders	Check Recommendation #3 Against Decision Matrix
9:45 A.M.	Break		

AGENDA
COUNCIL ON TEACHER EDUCATION

June 24-26, 1980
Wakulla Springs Inn

TIME	ITEM	WHO	ACTION REQUIRED OR TAKEN
Thursday, June 26, 1980 (continued)			
10:15 A.M.	General Session	~~Marian Calway~~	~~Action on Recommendation #3~~ *
		Annie Ward	Receive technical materials for Recommendation #4
11:30 A.M.	Lunch		
12:30 P.M.	Mixed Role Groups	Group Leaders	Modify Recommendation #4 and Rationale
1:30 P.M.	General Session	Dale Lake	Develop Second Draft of Recommendation #4
2:00 P.M.	Role - Alike Groups	Group Leaders	Check Recommendation #4 Against Decision Matrix
2:30 P.M.	General Session	Marian Calway	~~Action on Recommendation #4~~ *
3:30 P.M.	Adjourn		+ Action on all recommendations delayed until all were considered.

APPENDIX E

Forms

E.1 Contractors' Review Forms

E.2 Expert Panel Review Forms:
Content, Technical, Bias

APPENDIX E.1
ITEM REVIEW
Report Form

Reviewer's Perspective
(Check all which apply)

___ Classroom Teacher
___ Content area scholar and/or professor
___ Measurement Expert

Ethnic Group

___ Black ___ Hispanic
___ NonHispanic White ___ Other

Sex: ___ Male ___ Female

Competency _____
Subskill _____
Item No. (If more than 1 for subskill _____

(Circle One) Comment if NO or ?

1. Relevance to competency and subskill

   a. Does the item address the appropriate knowledge base for the competency and subskill? YES NO ?

   b. Is the knowledge base addressed correct and current? YES NO ?

2. Level of behavior

   a. Which level of behavior does this item require? YES NO ?

      Knowledge ___ Application ___
      Problem Solving ___

   b. Is the level of behavior appropriate for the competency and subskill? YES NO ?

3. Adequacy of item stem

   a. Does the stem make clear what is the task to be performed? YES NO ?

   b. Is the wording of the stem clear? YES NO ?

   c. Is the stem free of extraneous or confusing materials? YES NO ?

4. Relevance of scenario (if provided)

   a. Is scenario realistic? YES NO ?

   b. Is scenario appropriate for the competency and subskill? YES NO ?

	(Circle One)	Comment if NO or ?
5. Relevance of pictorial material (if provided)		
a. Is pictorial material appropriate for the competency and subskill?	YES NO ?	
b. Is pictorial material necessary for the item?	YES NO ?	
6. Appropriateness of item options		
a. What is the correct answer ___		
b. Is the correct answer stated accurately?	YES NO ?	
c. Are any other options also correct?	YES NO ?	
d. Are all the options grammatically parallel?	YES NO ?	
e. Are all the options logically parallel?	YES NO ?	
f. Are all options plausible for naive examinees?	YES NO ?	
g. Are all options free of irrelevant cues?	YES NO ?	
h. Are all options of approximately the same length?	YES NO ?	
i. Are the options reasonably ordered?	YES NO ?	
7. General item characteristics		
a. Does the item fit the item specifications?	YES NO ?	
b. Is the item consistent with required formatting?	YES NO ?	
c. Does the item have "face validity?"	YES NO ?	
d. Is the reading level appropriate?	YES NO ?	
8. Item Bias		
a. Does the item contain any information that could be seen as offensive to the culture, race or religion of teacher applicants?	YES NO ?	

	(Circle One)	Comment if NO or ?
b. Does the item contain any information that could be seen as offensive to either sexual group?	YES NO ?	
c. Does the item include stereotypic depictions of any cultural, racial, or religious group that are debasing?	YES NO ?	
d. Does the item include stereotypic depictions of either sexual group that are debasing?	YES NO ?	
e. Does the item portray cultural, racial, or religious groups as unequal in ability or natural endowment?	YES NO ?	
f. Does the item portray sexual groups as unequal in ability or natural endowments?	YES NO ?	
g. Does the item contain clues or information that could be seen to work to the benefit or detriment of any cultural, racial, or religious group?	YES NO ?	
h. Does the item contain clues or information that could be seen to work to the benefit or detriment of either sexual group?	YES NO ?	
i. Does the item contain group-specific language or vocabulary (e.g., culture-related expressions or slang)?	YES NO ?	

Form I-R-1

APPENDIX E.2
Panel Review Form

READING

Passage Code _____

Key:

1 _____
2 _____
3 _____
4 _____
5 _____
6 _____
7 _____
8 _____
9 _____
10 _____

Passage and Items

Appropriateness (How important is it that prospective teachers be able to read and understand material such as this?

1. Essential _____
2. Important _____
3. Acceptable _____
4. Questionable _____

Difficulty Level

1. Easy _____
2. Medium _____
3. Hard _____

Form I-M-1

Item Code _____

Key _____

TASK I MATHEMATICS

ITEM

Traceback

Competency _____
Subskill _____
Item
Descriptor _____

Appropriateness (How important is it that prospective teachers be able to answer this item correctly?

1. Essential _____
2. Important _____
3. Acceptable _____
4. Questionable _____

Difficulty Level

1. Easy _____
2. Medium _____
3. Hard _____

Form I-M-2

Item Code _____

MATHEMATICS

ITEM:

Appropriateness (How important is it that prospective teachers be able to answer this item correctly?

1. Essential _____
2. Important _____
3. Acceptable _____
4. Questionable _____

Difficulty Level

1. Easy _____
2. Medium _____
3. Hard _____

Form I-P-1

Item Code _____

PROFESSIONAL EDUCATION

ITEM

Correct Answer _____

Competency _____
Subskill _____
Content Base Category(ies):

I _____
II _____
III _____
IV _____
V _____
VI _____

Appropriateness (How important is it that teachers and teacher candidates be able to answer this item correctly?)

1. Essential _____
2. Important _____
3. Acceptable _____
4. Questionable _____

Difficulty Level:

1. Easy _____
2. Medium _____
3. Hard _____

Form I-P-2

Item Code _____

PROFESSIONAL EDUCATION

ITEM

Correct Answer _____

Appropriateness (How important it is that prospective teachers be able to answer this item correctly?

1. Essential _____
2. Important _____
3. Acceptable _____
4. Questionable _____

Difficulty Level:

1. Easy _____
2. Medium _____
3. Hard _____

RATING:
Accept _____
Minor Revisions _____
Major Revisions _____
Reject _____

CONTENT REVIEW - TASK II
MATHEMATICS

Item No. _____
Competency _____
Subskill _____

Reviewer Code _____

	Circle One	Comments
1. General item characteristic		
a. Does the item fit the item specifications?	YES NO ?	
b. Is the item consistent with required formatting?	YES NO ?	
c. Does the item have "face validity"?	YES NO ?	
d. Is the reading level appropriate?	YES NO ?	
2. Adequacy of item stem		
a. Does the stem make clear what is the task to be performed?	YES NO ?	
b. Is the wording of the stem clear?	YES NO ?	
c. Is the stem free of confusing materials?	YES NO ?	
d. Does the stem contain extraneous materials? (May be required by the specification.)	YES NO ?	
3. Pictorial Material (If used)		
a. Is the pictorial material appropriate for the subskill?	YES NO ?	
b. Is the pictorial material accurate?	YES NO ?	
4. Appropriateness of item options		
a. Is the style of the options correct and clear?	YES NO ?	
b. Are the options arranged in a logical and systematic order?	YES NO ?	
5. Behavior Required		
a. What is the level of behavior required by the item? (Check highest.)	Computation _____ Understanding _____	Problem Solving _____ Analysis/Synthesis _____
b. Can the item be answered with a "ball park" estimate?	YES NO	
c. How many steps are required to solve the problem?	1 2 3 More	
d. Is the item "Real World"?	YES NO	
e. If "Real World," is it (1) Teacher-Related? (2) Consumer-related?	YES NO YES NO	
6. Is this a measurement item? If Yes, answer "a" and "b" below:	YES NO	
a. Is it (1) English system? (2) Metric System?		
b. Which topic(s) is (are) addressed?		
(1) Length	_____	
(2) Weight	_____	
(3) Capacity	_____	
(4) Area/Perimeter/Volume	_____	
(5) Elapsed Time	_____	
7. Is "None of the above" used as an option? If Yes, answer "a" and "b" below:	YES NO	
a. Is "None of the above" the correct answer?	YES NO	
b. Is the correct answer a finite, exact value?	YES NO	

RATING
Accept ____
Minor Revisions ____
Major Revisions ____
Reject ____

Competency ____
Subskill ____
Sub-subskill ____
Item Number ____

Keyed Answer ____

Item Code ____

Reviewers Code ____

CONTENT REVIEW - TASK II
PROFESSIONAL EDUCATION

	Circle One	Comment if "NO" or "?"
1. Relevance to competency and subskill		
a. Does the item address the appropriate knowledge base for the competency and subskill?	YES NO ?	
b. Is the knowledge base addressed correct and current?	YES NO ?	
2. Level of behavior		
a. Which level of behavior does this item require?	Knowledge ____ Application ____ Problem Solving ____	
b. Is the level of behavior appropriate for the competency and subskill?	YES NO ?	
3. Adequacy of item stem		
a. Does the stem make clear what is the task to be performed?	YES NO	
b. Is the wording of the stem clear?	YES NO ?	
c. Is the stem free of extraneous or confusing materials?	YES NO ?	
4. Relevance of scenario (if provided)		
a. Is scenario realistic?	YES NO ?	
b. Is scenario appropriate for the competency and subskill?	YES NO ?	
5. Relevance of pictorial material (if provided)		
a. Is pictorial material appropriate for the competency and subskill?	YES NO ?	
b. Is pictorial material necessary for the item?	YES NO ?	
6. Appropriateness of item options		
a. Is the keyed answer stated accurately?	YES NO ?	
b. Are any other options also correct?	YES NO ?	
c. Are all the options grammatically parallel?	YES NO ?	
d. Are all the options logically parallel?	YES NO ?	
e. Are all options plausible for naive examinees?	YES NO ?	
f. Are all options free of irrelevant clues?	YES NO ?	
g. Are all options of approximately the same length?	YES NO ?	
h. Are the options reasonably ordered?	YES NO ?	
7. General item characteristics		
a. Does the item fit the item specifications?	YES NO ?	
b. Does the item have "face validity"?	YES NO ?	
c. Is the reading level appropriate?	YES NO ?	

RATING
Accept _____
Minor Revisions _____
Major Revisions _____
Reject _____

Topic Code: _____

Reviewers Code _____

TASK II
MEASUREMENT REVIEW
WRITING - 1

INSTRUCTIONS:	Circle One	Comments
a. Is the tone friendly and supportive?	YES NO ?	_____
b. Are the instructions clear about the basis for the scoring?	YES NO ?	_____
c. Are examinees encouraged to plan and organize their thoughts before writing?	YES NO ?	_____
TOPICS:		
a. Is the topic self-explanatory; that is, is it clearly and explicitly phrased?	YES NO ?	_____
b. Is the topic adequately defined and delimited?	YES NO ?	_____
c. Is the topic likely to be familiar to every examinee?	YES NO ?	_____
d. Is the topic stimulating; i.e., will it evoke examinee interest?	YES NO ?	_____
e. Is the topic fresh; i.e., is it one which has not been over-used?	YES NO ?	_____
f. Is the emotional tone appropriate, i.e., neither too pedestrian nor too sensational?	YES NO ?	_____

Form II-M-W-1

Topic Code _____

Reviewers Code _____

TASK II
MEASUREMENT REVIEW
WRITING - 2

In answering these questions, refer to the paper "Validation of Topics and Comparison of Three Presentation Modes for the Writing Subtest of the Florida Teacher Certification Examination."

	Circle One	Comments:
1. Topics		
a. Do any topics yield lower ratings than others?	YES NO ?	_____
b. Do any topics yield higher ratings than others?	YES NO ?	_____
c. Do any topics yield shorter essays than others?	YES NO ?	_____
d. Do any topics yield longer essays than others?	YES NO ?	_____
2. Modes		
a. Do any modes yield lower ratings than others?	YES NO ?	_____
b. Do any modes yield higher ratings than others?	YES NO ?	_____
c. Do any modes yield shorter essays than others?	YES NO ?	_____
d. Do any modes yield longer essays than others?	YES NO ?	_____
3. Interaction		
a. Does there appear to be an interaction of topic and mode?	YES NO ?	_____

4. Recommendations

a. Acceptable Topics: _____
_____

b. Recommended Mode: _____
_____

RATING
Accept _____
Minor Revisions _____
Major Revisions _____
Reject _____

Passage Code _____

Reviewers Code _____

TASK II
MEASUREMENT REVIEW
READING - 1

	Circle One	Comment if "NO" or "?"
1. Are the directions clear and easy to follow?	YES NO ?	
2. Are any distractors too close in meaning to the keyed answer?	YES NO ?	
3. Do the deleted words follow the specifications?	YES NO ?	
4. Are all the distractors for each item syntactically equivalent to the keyed answer?	YES NO ?	
5. Are all options for each item approximately equal in length?	YES NO ?	

RATING
Accept _____
Minor Revisions _____
Major Revisions _____
Reject _____

Passage Code _____

Reviewers Code _____

TASK II
MEASUREMENT REVIEW
READING - 2

	Circle One	Comment if "NO" or "?"
1. Appropriateness of Format		
a. Is the format of the item appropriate for the content and skill being tested?	YES NO ?	
2. Adequacy of Item Stem		
a. Does the stem make clear what is the task to be performed?	YES NO ?	
b. Is the wording of the stem clear?	YES NO ?	
c. Is the stem free of extraneous or confusing materials?	YES NO ?	
3. Item Data		
a. Is the difficulty level acceptable?	YES NO ?	
b. Is the discrimination level acceptable?	YES NO ?	
c. Are all distractors functioning?	YES NO ?	

Form II-M-R2

RATING
Accept _____
Minor Revisions _____
Major Revisions _____
Reject _____

Passage Code _____

Reviewers Code _____

TASK II
MEASUREMENT REVIEW
MATHEMATICS

	Circle One	Comment if "NO" or "?"
1. Appropriateness of Format		
a. Is the format of the item appropriate for the content and skill being tested?	YES NO ?	
2. Adequacy of Item Stem		
a. Does the stem make clear what is the task to be performed?	YES NO ?	
b. Is the wording of the stem clear?	YES NO ?	
c. Is the stem free of extraneous or confusing materials?	YES NO ?	
3. Item Data		
a. Is the difficulty level acceptable?	YES NO ?	
b. Is the discrimination level acceptable?	YES NO ?	
c. Are all distractors functioning?	YES NO ?	

RATING
Accept _____
Minor Revisions _____
Major Revisions _____
Reject _____

Item Code _____

Reviewers Code _____

TASK II
MEASUREMENT REVIEW
PROFESSIONAL EDUCATION - 1

	Circle One	Comments:
1. Appropriateness of Format		
a. Is the format of the item appropriate for the content and skill being tested?	YES NO ?	
2. Adequacy of Item Stem		
a. Does the stem make clear what is the task to be performed?	YES NO ?	
b. Is the wording of the stem clear?	YES NO ?	
c. Is the stem free of extraneous or confusing materials?	YES NO ?	
3. Appropriateness of Item Options		
a. Is the keyed answer stated accurately?	YES NO ?	
b. Are any other options also correct?	YES NO ?	
c. Are all the options grammatically parallel?	YES NO ?	
d. Are all the options logically parallel?	YES NO ?	
e. Are the options clearly worded?	YES NO ?	
f. Are all options plausible for naive examinees?	YES NO ?	
g. Are all options free of irrelevant clues?	YES NO ?	
h. Are all options of approximately the same length?	YES NO ?	
i. Are the options reasonably ordered?	YES NO ?	
4. Language Usage		
a. Does the language of this item follow standard usage?	YES NO ?	
b. Is the item free of ambiguities?	YES NO ?	
c. Are all the referents explicit and correct?	YES NO ?	
d. Is the item free of misplaced modifier phrases or clauses?	YES NO ?	

Form II M-P-1

RATING
Accept _____
Minor Revisions _____
Major Revisions _____
Reject _____

Item Code _____

Reviewers Code _____

TASK II
MEASUREMENT REVIEW
PROFESSIONAL EDUCATION - 2

	Circle One	Comments:
1. Appropriateness of Format		
a. Is the format of the item appropriate for the content and skill being tested?	YES NO ?	
2. Adequacy of Item Stem		
a. Does the stem make clear what is the task to be performed?	YES NO ?	
b. Is the wording of the stem clear?	YES NO ?	
c. Is the stem free of extraneous or confusing materials?	YES NO ?	
3. Appropriateness of Item Options		
a. Is the keyed answer stated accurately?	YES NO ?	
b. Are any other options also correct?	YES NO ?	
c. Are all the options grammatically parallel?	YES NO ?	
d. Are all the options logically parallel?	YES NO ?	
e. Are the options clearly worded?	YES NO ?	
f. Are all options plausible for naive examinees?	YES NO ?	
g. Are all options free of irrelevant clues?	YES NO ?	
h. Are all options of approximately the same length?	YES NO ?	
i. Are the options reasonably ordered?	YES NO ?	
4. Item Data		
a. Is the difficulty level acceptable?	YES NO ?	
b. Is the discrimination level acceptable?	YES NO ?	
c. Are all distractors functioning?	YES NO ?	

Form II M-P-2

RATING
Accept _____
Minor Revisions _____
Major Revisions _____
Reject _____

Passage No. _____

Reviewers Code _____

TASK II
BIAS REVIEW
READING

	Circle One	Comment if "YES" or "?"
1. Does the passage contain any information that could be seen as offensive to the culture, race, or religion of teacher applicants?	YES NO ?	_____
2. Does the passage contain any information that could be seen as offensive to either sexual group?	YES NO ?	_____
3. Does the passage include stereotypic depictions of any cultural, racial, or religious group that are debasing?	YES NO ?	_____
4. Does the passage include stereotypic depictions of either sexual group that are debasing?	YES NO ?	_____
5. Does the passage portray cultural, racial, or religious groups as unequal in ability or natural endowments?	YES NO ?	_____
6. Does the passage portray sexual groups as unequal in ability or natural endowments?	YES NO ?	_____
7. Does the passage contain clues or information that could be seen to work to the benefit or detriment of any cultural, racial, or religious group?	YES NO ?	_____
8. Does the passage contain clues or information that could be seen to work to the benefit or detriment of either sexual group?	YES NO ?	_____
9. Does the passage contain group-specific language or vocabulary (e.g., culture-related expressions or slang)?	YES NO ?	_____

RATING
Accept _____
Minor Revisions _____
Major Revisions _____
Reject _____

Topic Code _____

Reviewers Code _____

TASK II
BIAS REVIEW
WRITING

	Circle One	Comment if "YES" or "?"
1. Does the item contain any information that could be seen as offensive to the culture, race or religion of teacher applicants?	YES NO ?	
2. Does the item contain any information that could be seen as offensive to either sexual group?	YES NO ?	
3. Does the item include stereotypic depictions of any cultural, racial, or religious group that are debasing?	YES NO ?	
4. Does the item include stereotypic depictions of either sexual group that are debasing?	YES NO ?	
5. Does the item portray cultural, racial, or religious groups as unequal in ability or natural endowments?	YES NO ?	
6. Does the item portray sexual groups as unequal in ability or natural endowments?	YES NO ?	
7. Does the item contain clues or require information that could be seen to work to the benefit or detriment of any cultural, racial, or religious group?	YES NO ?	
8. Does the item contain clues or require information that could be seen to work to the benefit or detriment of either sexual group?	YES NO ?	
9. Does the item contain group-specific language or vocabulary (e.g., culture-related expressions or slang)?	YES NO ?	

RATING
Accept _____
Minor Revisions _____
Major Revisions _____
Reject _____

Item Code:

Reviewers Code _____

TASK II
BIAS REVIEW
MATHEMATICS, PROFESSIONAL EDUCATION

	Circle One	Comment if "YES" or "?"
1. Does the item contain any information that could be seen as offensive to the culture, race, or religion of teacher applicants?	YES NO ?	
2. Does the item contain any information that could be seen as offensive to either sexual group?	YES NO ?	
3. Does the item include stereotypic depictions of any cultural, racial, or religious group that are debasing?	YES NO ?	
4. Does the item include stereotypic depictions of either sexual group that are debasing?	YES NO ?	
5. Does the item portray cultural, racial, or religious groups as unequal in ability or natural endowments?	YES NO ?	
6. Does the item portray sexual groups as unequal in ability or natural endowments?	YES NO ?	
7. Does the item contain clues or information that could be seen to work to the benefit or detriment of any cultural, racial, or religious group?	YES NO ?	
8. Does the item contain clues or information that could be seen to work to the benefit or the detriment of either sexual group?	YES NO ?	
9. Does the item contain group-specific language or vocabulary (e.g., culture-related expressions or slang)?	YES NO ?	

Form II-B-M-P

APPENDIX F

TECHNICAL MATERIALS AND INSTRUCTIONS USED BY STANDARD-SETTING PANEL

TENTATIVE RECOMMENDATIONS

COTE
June 24-26, 1980
Wakulla Springs, Florida

A recommendation of the performance level which is to be considered "passing" for each of the subtests is provided for consideration. Technical information will be provided to help you consider and modify each recommendation.

In addition to these materials, you will have samples of calibrated items for the three multiple-choice tests and writing samples scored at various levels. These materials will help you get a feeling for the performance level represented by the various scores.

Descriptive data about the field test samples is provided in the attached table.

Tentative Recommendations

Recommendation #1 - It is recommended that the cutting score for the Reading Test be set at a point which represents 75% comprehension on a sample of materials from the identified and adopted domain. This is a logit value of 1.4 which will be converted to an FTCE scale value of 200. (This is close to 80% comprehension and fits the data.)

Recommendation #2 - The minimal acceptable performance level on the Mathematics Subtest should be that equal to a logit value of 1.00.

Recommendation #3 - The minimal acceptable performance level on the Professional Education Subtest should be that equal to a logit value of .25.

Recommendation #4 - An individual shall pass the Writing Subtest if at least two of three trained judges determine that the writing sample is acceptable; that is, has a total score of 5 or more. Thus a minimum passing score, based on the sum of three judges' ratings, will be 5.

INSTRUCTIONS FOR READING ITEMS

You have a sample of passages to be used for the Reading subtest, arranged in order from easy to hard. The difficulty level is indicated, with -1.88 being easy and 3.29 being difficult.

It is suggested that you follow these steps:

(1) "Take" the items as a test.

(2) Check your answers against the scoring key.

(3) Sort the passages into two piles:

A. Applicants should be expected to answer most items correctly (all but 1 or 2).

B. Applicants may miss several items in each passage (3 or more).

(4) Determine the dividing point (in logits) between the two groups.

(5) Compare your cutting point with that of others in your group and arrive at a group decision.

(6) Compare the group decision with Tentative Recommendation #1.

Recommendation #1 - Reading
Technical Materials

Domain of Materials Sampled

The domain of materials which are sampled for the Reading Subtest were defined as follows:

A. Textbooks: Textbooks for commonly required professional education courses. These were identified by a survey of all Florida institutions with approved teacher education programs. The textbooks were organized into these categories:

1. History, Philosophy, Social Foundations, School Law
2. Methods
3. Measurement
4. Educational Psychology

B. Library Journals: Non-specialized educational journals.

C. State Department of Education Publications: Selected Department of Education publications of general interest.

Testing Technique

Multiple-choice cloze test. Test items constructed by randomly selecting passages from the domain, deleting each nth word and providing 4 options from which examinee is to select the correct word.

Research Evidence:

Refer to article by John R. Bormuth, "Empirical Determination of the Instructional Reading Level," International Reading Association Conference Proceedings, 1969, 13, 716-721.

Summary:

Reading experts have traditionally used a criterion of 75% mastery to determine the "instructional level" of reading materials. This practice is based on a recommendation by Thorndike (1917) and widely promulgated by Betts (1946), although there has been little empirical evidence to support the practice.

Bormuth conducted a study to investigate the matter. The specific question addressed was whether "...there is some range of difficulty which maximizes the amount of information students gain as a consequence of reading instructional materials."

The results of this study supported the use of 75% comprehension as being optimal for learning gain.

Interpretation of the data is complicated by the fact that Bormuth use completion CLOZE tests and translated his results into multiple-choice equivalents, based on another study.

*article attached

Normative Information

The distribution of scores for the field-test sample is graphed on the attachment. Setting the cutting point at a logit of 1.4, which is equivalent to 80% comprehension, would result in 6.6% failing this subtest.

READING SUBTEST (50 Items)
DISTRIBUTION OF SCORES
FIELD TEST - April 1980

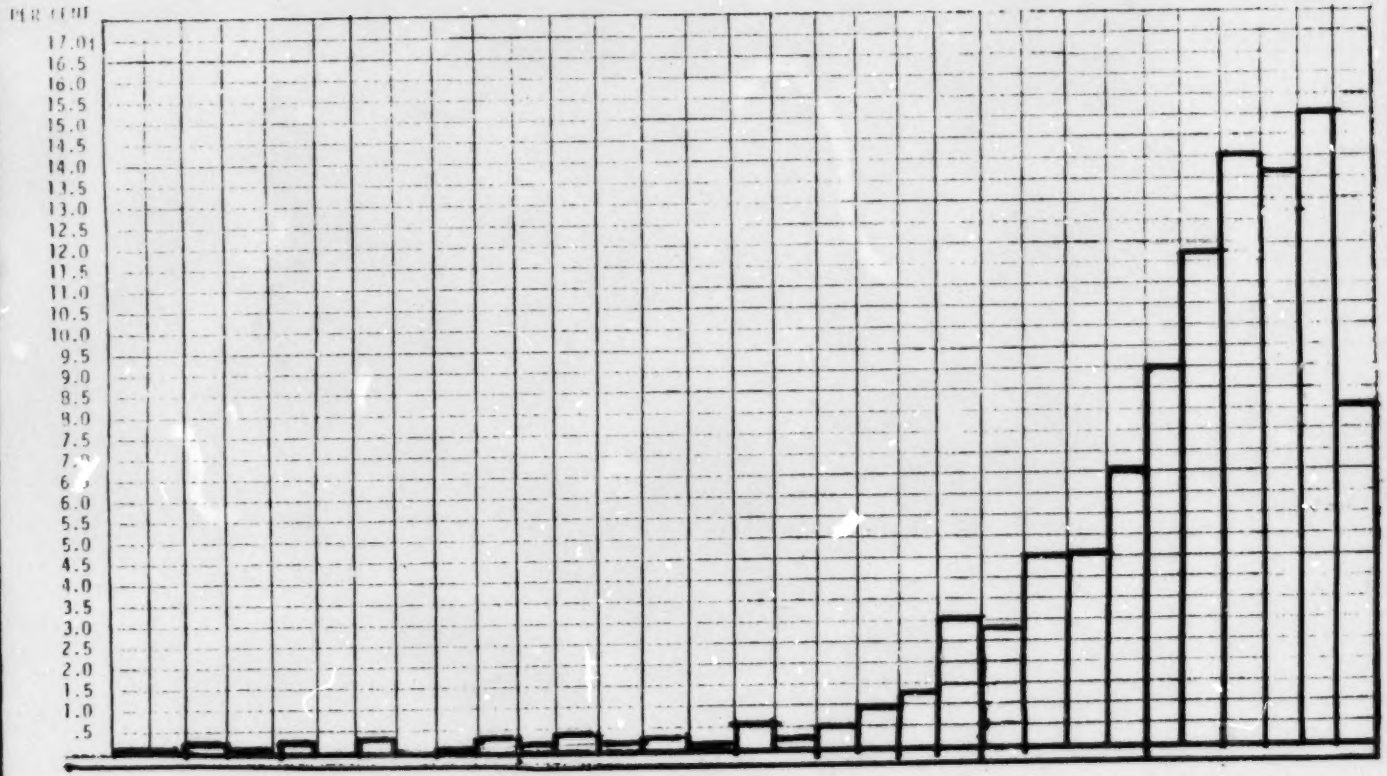
		STATEWIDE		
RAW SCORE	LOGIT*	N	%	CUM. %
20 & Below	- .40	1	.08	.08
21	- .30	1	.08	.16
22	- .20	2	.17	.33
23	- .15	1	.08	.41
24	- .10	2	.17	.58
25	.00			.58
26	.10	4	.34	.92
27	.15			.92
28	.20	1	.08	1.00
29	.30	4	.34	1.34
30	.40	2	.17	1.51
31	.50	5	.42	1.93
32	.60	3	.25	2.18
33	.70	5	.42	2.60
34	.75	4	.34	2.94
35	.80	7	.59	3.53
36	.90	4	.34	3.87
37	1.00	6	.51	4.38
38	1.10	11	.93	5.31
39	1.20	15	1.26	6.57
40	1.40	37	3.12	9.69
41	1.50	32	2.70	12.39
42	1.65	54	4.55	16.94
43	1.80	55	4.64	21.58
44	2.00	79	6.66	28.24
45	2.20	107	9.02	37.26
46	2.40	139	11.72	48.98
47	2.80	167	14.08	63.06
48	3.20	162	13.66	76.72
49	3.60	180	15.18	91.90
50	4.00 +	96	8.09	99.99
		1,186		

*From Rasch Analysis

READING

GROUP STATEWIDE

N 1186



INSTRUCTIONS FOR MATHEMATICS ITEMS

You have a sample of mathematics items, arranged in order from easy to hard. The difficulty level is indicated, with -1.30 being easy and 1.50 being difficult.

It is suggested that you follow these steps:

(1) "Take" the items as a test.

(2) Check your answers against the scoring key.

(3) Sort the items into two piles:

A. Applicants should be expected to answer correctly more than half of items of this type.

B. Applicants should be expected to answer correctly less than half of items of this type.

(4) Determine the dividing point (in logits) between the two groups.

(5) Compare your cutting point with that of others in your group and arrive at a group decision.

(6) Compare the group decision with Tentative Recommendation #2.

Recommendation #2 - Mathematics
Technical Materials

Description of Test Items

The examination covers only basic mathematics, with very little Algebra or other higher mathematics. Half the items are simple computation or one-step word problems, the other half are "real world" problems.

The items are multiple choice, and only 36% include "None of the Above" as an option.

In general, the items are judged to be fairly "easy."

Normative Information

The distribution of field-test scores is quite skewed. For the 27 items, the mean was approximately 22; SD was approximately 5. That means that 2/3 of the scores fell between 17 and 27. The logit for 17 is .50. The "break" in the distribution comes between 19 and 20, with a logit value of 1.00.

Setting the cutting score at .90 will result in approximately 30% failing. A cutting score of .50 would fail 16%.

Additional Information

The "passing" score for student assessment for similar items is 70%. For 27 items, that would be a raw score of 19, with a logit value of .80.

MATHEMATICS SUBTEST (27 Items)
DISTRIBUTION OF SCORES
FIELD TEST - April 1980

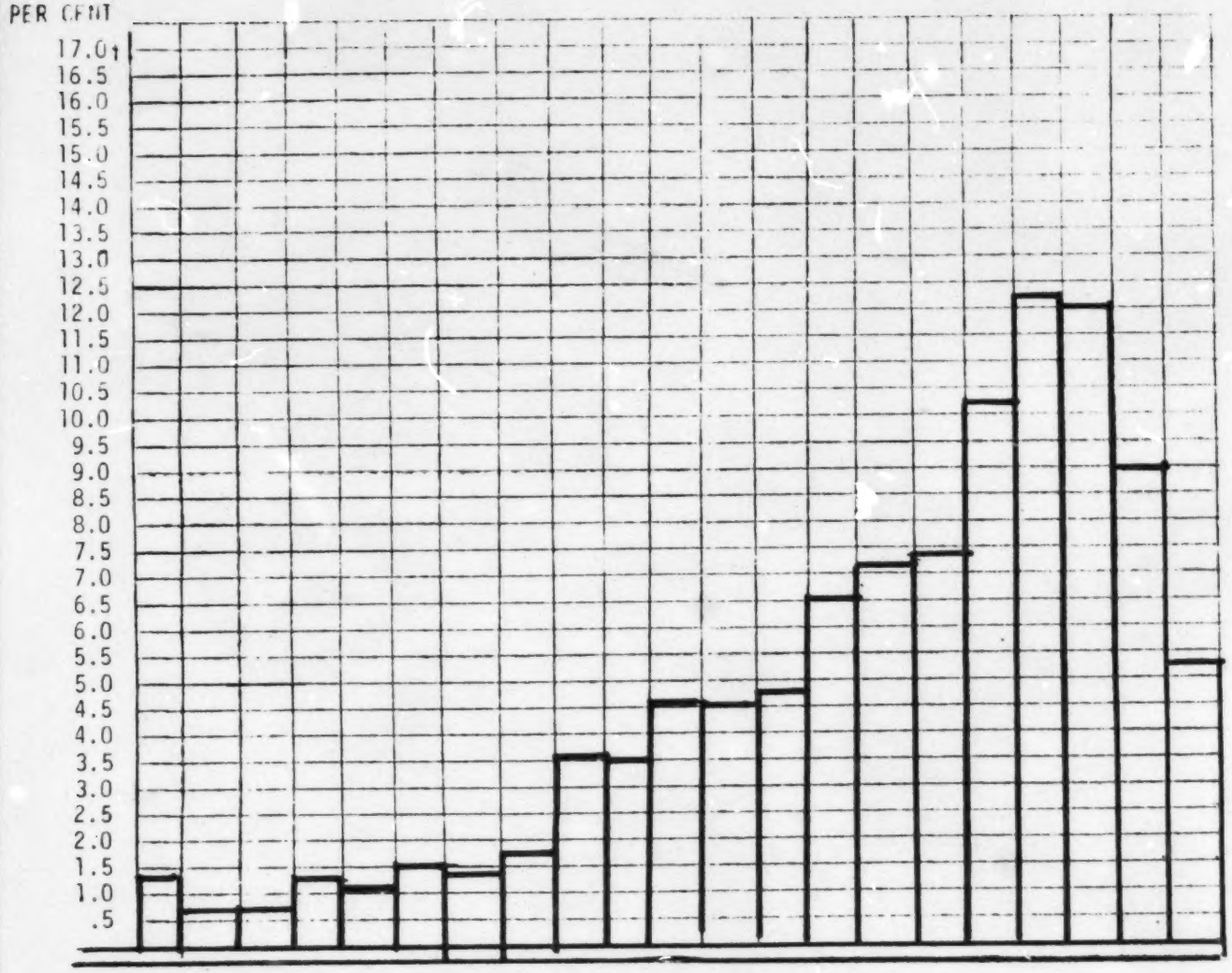
RAW SCORE	LOGIT*	STATEWIDE N	%	CUM %
7 & Below	-1.40	16	1.4	1.4
8	-1.20	8	.7	2.1
9	-1.00	8	.7	2.8
10	- .80	16	1.4	4.2
11	- .65	13	1.1	5.3
12	- .40	18	1.5	6.8
13	- .20	17	1.4	8.2
14	- .10	21	1.8	10.0
15	.10	37	3.1	13.1
16	.30	36	3.0	16.1
17	.50	54	4.6	20.7
18	.60	53	4.5	25.2
19	.80	57	4.8	30.0
20	1.00	78	6.6	36.6
21	1.30	87	7.3	43.9
22	1.50	88	7.4	51.3
23	1.70	122	10.3	61.6
24	2.00	145	12.2	73.8
25	2.40	142	12.0	85.8
26	2.90	106	8.9	94.7
27	3.60	64	5.3	100.0
		1,186		

* From Rasch Analysis

GROUP STATEWIDE

N- 1186

MATHEMATICS



RAW SCORE	below 7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
LOGIT	-1.40	-1.20	-1.00	-.80	-.65	-.40	-.20	-.10	.10	.30	.50	.60	.80	1.00	1.30	1.50	1.70	2.00	2.40	2.90	3.60

Recommendation #3
Technical Materials

Description of the Test Item

a. The test will consist of 100 multiple-choice items.

b. The test is generic; that is, it covers general principles and recommended procedures, rather than those related to specialized areas.

c. Most of the items have been written at the application or problem-solving level. This was done in order to provide a test task which is as close to actual classroom situations as possible with a paper and pencil test.

d. In writing application and problem-solving items, a scenario is frequently used to present a context which specifies important, relevant variables. There are two potential problems associated with this practice:

   1. Some degree of ambiguity may result since it is not always possible to specify all relevant variables.

   2. The context chosen may not be familiar to a given examinee.

The multiple review processes have addressed the first problem to some extent; the second has been handled by deliberately varying the context so that there is a balance between various grade levels and subjects. This means that every examinee will have some items for which the context may be unfamiliar, even though the principle being tested is one which should be familiar to all teacher applicants. (See attached analysis.)

The correctness of the keyed response can be established either by reference to professional literature or by expert opinion or both. Experts differ to some extent as to the correctness of the keying of the items. Application and problem-solving require the integration of many issues and theories, so the correct answer is seldom clear-cut. The keyed answer is the one selected most frequently by "experts," but the percent of experts choosing the keyed answer varies from one item to another.

Normative Information

Except for a small percentage of cases with scores at or near the chance level, score on the Professional Education Subtest are almost normally distributed, with a mean of approximately 34, SD of approximately 6. The modal score is 36, with 38 being the next most frequent score.

The distribution has a slight "break" between a logit of .40 and .45. Setting the cutting score at .45 would fail 20%.

Raw scores lower than 25, made by 9% of the sample, probably have a large component of chance.

The recommended cutting point of .25 lies at -1 SD from the mean and above the 50% chance level. The percent failing with this cutting score would be 13%.

INSTRUCTIONS FOR PROFESSIONAL EDUCATION ITEMS

You have a sample of items for the Professional Education subtest, arranged in order from easy to hard. The difficulty level is indicated, with -1.60 being easy and 2.10 being difficult.

It is suggested that you follow these steps:

(1) "Take" the items as a test.

(2) Check your answers against the scoring key.

(3) Sort the items into two piles:

A. Applicants should be expected to answer correctly more than half of items of this type.

B. Applicants should be expected to answer correctly less than half of items of this type.

(4) Determine the dividing point (in logits) between the two groups.

(5) Compare your cutting point with that of others in your group and arrive at a group decision.

(6) Compare the group decision with Tentative Recommendation #3.

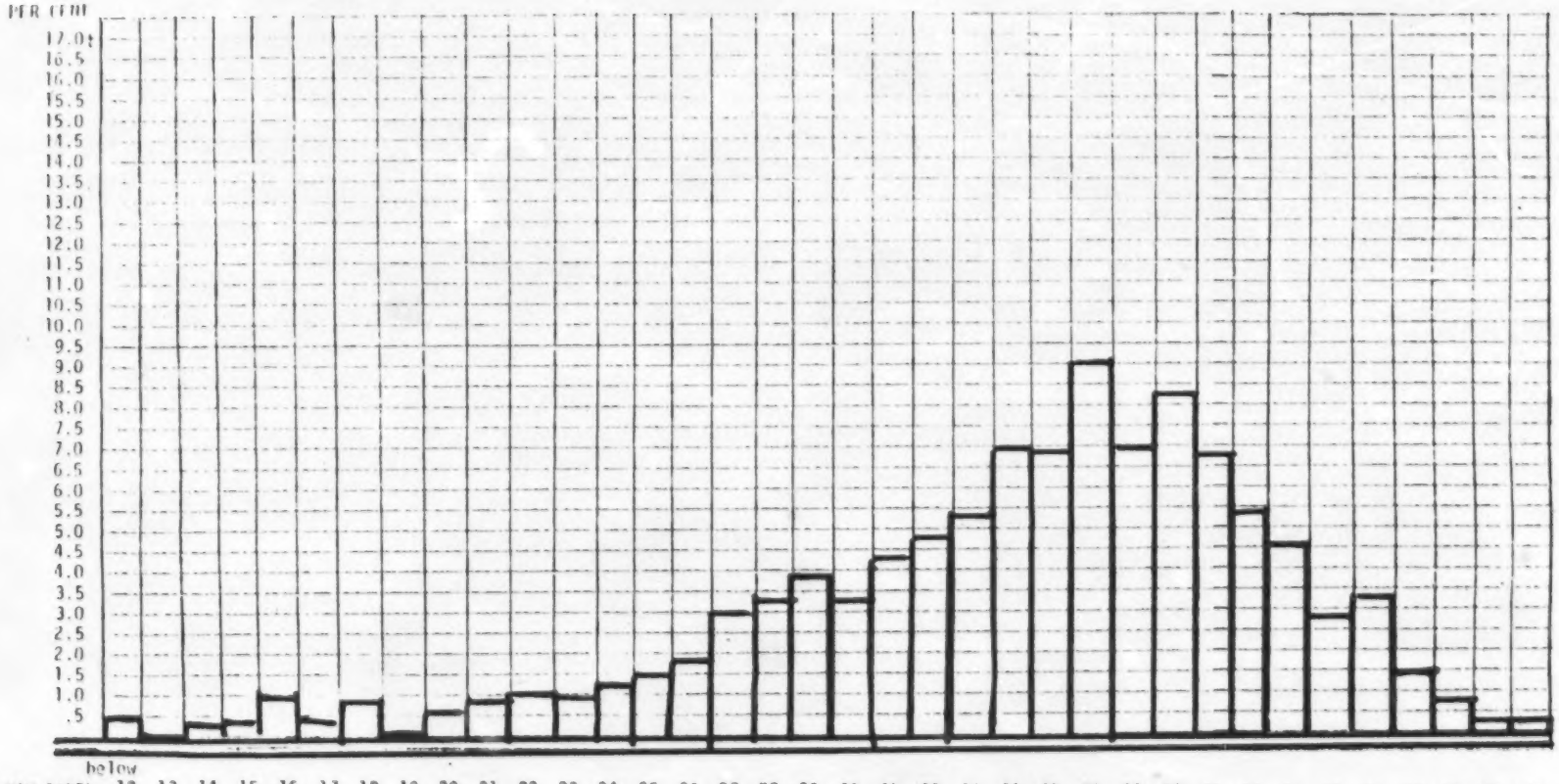
PROFESSIONAL EDUCATION SUBTEST
DISTRIBUTION OF SCORES
FIELD TEST - April 1980

RAW SCORE	LOGIT*	STATEWIDE N	%	CUM. %
12 & Below	-1.35	6	.5	.5
13	-1.27	1	.1	.6
14	-1.10	3	.3	.9
15	-1.00	4	.4	1.2
16	- .95	10	.9	2.1
17	- .80	4	.4	2.5
18	- .65	8	.7	3.2
19	- .60	2	.2	3.3
20	- .50	7	.6	4.0
21	- .40	9	.8	4.7
22	- .30	11	1.0	5.7
23	- .20	10	.9	6.5
24	- .16	14	1.2	7.8
25	.00	17	1.5	9.3
26	.05	19	1.7	11.0
27	.20	27	2.4	13.3
28	.25	31	2.7	16.1
29	.40	43	3.8	19.8
30	.45	36	3.2	23.0
31	.50	55	4.8	27.8
32	.70	59	5.2	33.0
33	.80	66	5.8	38.8
34	.92	79	6.9	45.7
35	1.00	78	6.8	62.5
36	1.10	104	9.1	61.6
37	1.20	73	6.4	68.0
38	1.30	93	8.2	76.2
39	1.50	72	6.3	82.5
40	1.60	61	5.3	87.8
41	1.70	46	4.0	91.9
42	2.00	26	2.3	94.1
43	2.10	32	2.8	96.9
44	2.40	20	1.4	96.7
45	2.60	9	.8	99.5
46	2.70	3	.3	99.7
47	3.20	3	.3	100.0
		1,141		

*From Rasch Analysis

GROUP STATEWIDE
N= 1141

PROFESSIONAL EDUCATION



DIRECTIONS FOR WRITING SAMPLES

You have eleven writing samples, selected to represent a wide range of scores assigned by trained judges. Some are "Unacceptable" (score of 3), some are "Minimally acceptable" (score of 5 or 6), others are "Acceptable" (score of 7-9) or "Superior" (score of 10-12)

It is suggested that you follow these steps:

(1) Sort the essays into 4 piles, according to your judgment as to overall quality, using the scoring criteria presented in Attachment 1. Your groups should be:

Unacceptable	(Rating of 1)
Minimally acceptable	(Rating of 2)
Acceptable	(Rating of 3)
Superior	(Rating of 4)

(2) Check your sorting against the judges' scores, using Table w-1.

(3) Consider whether the scoring directions, as written in Attachment 1, correctly classify the essays.

Recommendation #4 - Writing
Technical Materials

Description of the Test

The Writing Subtest will be a production writing sample scored by three trained and experienced judges, with discrepancies resolved by a "referee."

The standard for passing is built into the scoring standard. (See attachment 1.)

Preliminary studies have indicated that judges can be trained to attain very high reliabilities - better than 90% agreement.

Normative Information

The Writing Subtest was not included in the field test. However, an earlier study was carried out which provided some data. In the fall of 1979, writing samples were collected from 360 students at two teacher training institutions, Florida State University and University of South Florida. The sample is not completely representative of the population because certain ethnic and program groups were under-represented. The description of the sample and the distribution of scores is attached.

Because of the non-representativeness of the sample, it should be anticipated that the percent failing the Writing test will be higher than the 10% found in the study, perhaps being as high as 20%.

Scoring Procedure Recommended for the Writing Test by the Contractor

Scoring System to be Used

The scoring system for evaluating essays written on the Florida Teacher Certification Examination is as follows: each essay will be read independently by three raters, who will assign it a rating of 1, 2, 3, or 4, depending on his or her judgement of its overall quality. (The meaning of these ratings is discussed later.) If one of the three ratings differs from any other by two points or more, or if the three ratings are some combination of 1 1 2, the paper will be sent to a fourth rater, or referee, who will read it and assign it a rating. (This rating will then replace the most discrepant one.* The sum of three ratings for each paper--either the three original ones, or two originals and a referee's substitute--is its final score. The range of scores for each essay is thus 3 at the low end to 12 at the top.) Note that no paper will have a final score of 4. A minimally passing score is 5.

*If the referee's rating is 1, then it replaces the 2, so the score becomes 3. If it is 2 or higher, it replaces a 1, so the score becomes 5 or higher.

Description of Sample for Writing Study

FSU		USF	
Major	# in Sample	Major	# in Sample
Physical Ed.	55	Elementary Ed.	98
Special Ed.	21	Early Childhood Ed.	36
Music Ed.	17	Learning Disabilities	16
Elementary Ed.	17	EMR	8
English Ed.	13	EMH	6
Speech Pathology	10	English Ed.	2
Social Studies Ed.	7	Gifted Ed.	2
Home Economics Ed.	6	Foreign Language Ed.	1
Art Ed.	5	Deaf Ed.	1
Mathematics Ed.	5		170
Social Work	5		
Early Childhood Ed.	4		
Child Development	4		
Science Ed.	3		
Visual Disabilities	3		
Library Science	3		
Vocational/Business Ed.	2		
Industrial Arts Ed.	2		
Foreign Language Ed.	2		
Theater Ed.	1		
Political Science	1		
ESL	1		
Psychology	1		
Art Therapy	1		
Career Ed.	1		
	190		

Distribution of Scores on Writing Sample

Score	N	% of N
3	37	10.3
*4	0	0
5	100	27.8
6	71	19.7
7	76	21.1
8	47	13.1
9	13	3.6
10	13	3.6
11	2	.6
12	1	.3
Total	360	100

Mean Score--6.19
Median Score--6
Modal Score--5

*Papers with a score of four were reviewed and the scores were adjusted to either a three or a five depending on the referee's rating.

The Rating Criteria for the Writing Subtest

The criteria were developed to accommodate three specific conditions imposed by the writing test:

1. that they reflect those characteristics which are widely a[illegible]ed as indicative of good writing and which can thus reasonably be expected to [illegible]ppear in essays written by people seeking initial Florida Teacher Certification;

2. that they are amenable to being "translated" into operational descriptions of levels of competence that can be used by raters;

3. that they take into account the writing subskills identified by the Council on Teacher Education.

Criteria to be Used by Judges for the Evaluation of Essays

1. Rhetorical Quality

   1.1 Unity: An ordering and interdependence of parts producing a single effect: completeness.

   1.2 Focus: Concentration on the chosen topic.

   1.3 Clarity: Lucidity of expression; lack of ambiguity and distortion.

   1.4 Sufficiency: Appropriate depth and breadth of expression to meet the writer's purposes and the demands of the particular topic.

2. Structural and Mechanical Quality

   2.1 Organization: Consistent and coherent integration and connection of parts.

   2.2 Development: Appropriate and sufficient exposition of ideas; use of detail, examples, illustrations, comparisons, etc.

   2.3 Paragraph and Sentence Structure: Appropriate form, variety, logic, relatedness of and among structural units.

   2.4 Syntax: Appropriate ordering of words to convey intended meaning.

3. Observance of Conventions in Writing

   3.1 Usage: Appropriate [illegible]se of language features: inflections, tense, agreement, pronouns, modifiers, vocabulary, level of discourse, etc.

   3.2 Spelling, Capitalization, Punctuation: Consistent practice of accepted forms.

THE RATING SCALE

The four-level scale is used because it provides enough degrees of distinction to be meaningful in assessing writing competence.

1. The essay lacks unity and focus. It is distorted and/or ambiguous, and it fails to treat the topic in sufficient depth and breadth. There is little or no discernible organization and only scant development of ideas, if any at all. The essay betrays only sporadically a sense of paragraph and sentence structure, and it is syntactically slipshod. Usage is irregular and often questionable or wrong. There are serious errors in spelling, capitalization, and punctuation.

2. The essay has some degree of unity and focus, but each could be improved. It is reasonably clear, though not invariably so, and it treats the topic with a marginal degree of sufficiency. The essay reflects some concern for organization and some for development of ideas, but neither is necessarily consistent nor fully realized. The essay reveals some sense, if not full command, of paragraph and sentence structure. It is syntactically bland and at times, awkward. Usage is generally accurate, if not consistently so. There are some errors in spelling, capitalization, and punctuation that detract from the essay's effect if not from its sense.

3. The essay is focussed and unified, and it is clearly if not distinctively written. It gives the topic an adequate though not always thorough treatment. The essay is well organized, and must of the time it develops ideas appropriately and sufficiently. It shows a good grasp of paragraph and sentence structure, and its usage is generally accurate and sensible. Syntactically, it is clear and reliable. There may be a few errors in spelling, capitalization, and punctuation, but they are not serious.

4. The essay is unified, sharply focussed, and distinctively effective. It treats the topic clearly, completely, and in suitable depth and breadth. It is clearly and fully organized, and it develops ideas with consistent appropriateness and thoroughness. The essay reveals an unquestionably firm command of paragraph and sentence structure. Syntactically, it is smooth and often elegant. Usage is uniformly sensible, accurate, and sure. There are very few, if any, errors in spelling, capitalization, and punctuation.

APPENDIX G

References

Selected References

I. Professional Issues

Berliner, D. and Tikunoff, W. The California Beginning Teacher Evaluation Study: Overview of the Ethnographic Study. Journal of Teacher Education, Spring, 1976

Gage, N. L., Ed., Handbook of Research on Teaching, Chicago: Rand McNally, 1973

McNeil, J. D. and Popham, W. J. The Assessment of Teacher Competence, Second Handbook of Research on Teaching, R. M. Travers, Ed., American Educational Research Association, Chicago: Rand McNally, 1973

Medley, D. M. Teacher Competence and Teacher Effectiveness: A Review of Process - Product Research. Washington, D.C.: American Association of Colleges for Teacher Education, 1977

Peterson, D. and Ward, A., Ed., Due Process in Teacher Evaluation. Washington, D.C.: University Press of America, 1980

Popham, W. J. Performance Tests of Teaching Proficiency: Rationale, Development, and Validation, American Educational Research Journal, June, 1971

Rosenshine, B. and Furst, N. Research on Teacher Performance Criteria in Research in Teacher Education: A Symposium, B. Othanel Smith, Ed., Englewood Cliffs, NJ: Prentice-Hall, 1971

II. General Technical Issues

Bloom, B. S. Learning for Mastery, Handbook on Formative and Summative Evaluation of Student Learning. B. S. Bloom, J. T. Hastings, and G. F. Madaus, Eds., New York: McGraw-Hill, 1971

Hambleton, R. K. and Eignor, D. R. Guidelines for Evaluating Criteria-Referenced Tests and Test Manuals. Journal of Educational Research, 1978, 15

Hambleton, R. K. and Novick, M. R. Toward an Integration of Theory and Methods for Criterion-Referenced Tests. Journal of Educational Measurement, 1973, 10, 159-170

Standards for Educational and Psychological Tests. Washington, D.C.: American Psychological Association, 1974

Thorndike, R. L. (Ed.) Educational Measurement, Second Ed., Washington, D.C.: American Council on Education, 1971

III. Validity

Cronbach, L. J. Test Validation. Educational Measurement, R. L. Thorndike, Ed. Washington, D.C.: American Council on Education, 1971

Linn, R. L. Issues of Validity in Measurement for Competency-Based Programs, Practices and Procedures in Competency-Based Measurement. Washington, D.C.: NCME, 1979

IV. Reliability

Huynh, H. On the Reliability of Decisions in Domain-Referenced Testing. Journal of Educational Measurement, 1976, 13, 253-146

Linn, R. L. Issues of Reliability in Measurement for Competency-Based Programs. Practices in Problems in Competency-Based Measurement. Bunda and Sanders, Eds., Washington, D.C.: NCME, 1978

Stanley, J. C. Reliability. Educational Measurement, Second Edition, R. L. Thorndike, Ed., Washington, D.C.: American Council on Education, 1971

V. Bias

Angoff, W. H. A Technique for the Investigation of Cultural Differences. Paper - September, 1972

Angoff, W. H. and Ford, S. E. Item - Race Interaction on a Test of Scholastic Aptitude. Journal of Educational Measurement, Vol. 10, No. 2, Summer, 1973

Green, D. R. Racial and Ethnic Bias in Test Construction. CTB/McGraw-Hill, 1972

Harms, R. M. A Comparative Concurrent Validation of Selected Estimations of Test Items Bias. Doctoral Dissertation, University of South Florida, 1978

Hills, J. R. Cultural Bias in Minimum Competency Testing. Paper prepared for DOE, October 5, 1978

Holtzman, W. H. Validity and Legality. Educational Measurement and the Law, (Preceedings of the 1977 ETS Invitational Conference) Princeton, NJ, ETS, 1977

Lerner, B. Equal Protection and External Screening: Davis, Defunis, and Bakke, Educational Measurement and the Law, (Proceedings of the 1977 ETS Invitational Conference) Princeton, NJ, ETS, 1977

On Bias in Selection (Special Issue) Journal of Educational Measurement, R. M. Jaeger, Ed., Vol. 13, No. 1, Spring, 1976

Scheuneman, J. A Method of Assessing Bias in Test Items, Journal of Educational Measurement, Vol. 16, No. 3, Fall 1979

Shepard, L., Camilli, G. and Averill, M. Comparison of Six Procedures for Detecting Test Item Bias Using Both Internal and External Ability Criteria paper presented at NCME, Boston, April, 1980

Test Bias: A Bibliography. Eric Clearinghouse on Tests, Measurement, and Evaluation, June, 1971

VI. Score Scaling and Standard Setting

Angoff, William Scales, Norms and Equivalent Scores, in Educational Measurement (Second Ed.) R. L. Thorndike, Ed., Washington: American Council on Education, 1971

Conaway, L. E. Setting Standards in Competency-Based Education: Some Practices and Concerns, Practices and Problems in Competency-Based Measurement. M. A. Bunda and J. R. Sanders, Eds., Washington, D.C.: National Council on Measurement in Education, 1979

Ebel, R. L. Essentials of Educational Measurement, Englewood Cliffs, NJ: Prentice Hall, 1972

Emrich, J. A. An Evaluation Model for Mastery Testing, Journal of Educational Measurement, 1971, 8, 321-326

Glass, G. V. Standards and Criteria. Ms., University of Colorado, Date Unknown

Jaeger, R. M. Measurement Consequences of Selected Standard-Setting Models. Practices and Problems in Competency-Based Measurement. M. A. Bunda and J. R. Sanders, Eds., Washington, D.C.: National Council on Measurement in Education, 1979

Koffler, S. L. A Comparison of Approaches for Setting Proficiency Standards. Journal of Educational Measurement, Vol. 17, No. 3, Fall, 1980

Nedelsky, L. Absolute Grading Standards for Objective Tests. Educational and Psychological Measurement, 1954, 14, 3-19

Shepard, L. A. Setting Standards, Practicies and Problems in Competency-Based Measurement. M. A. Bunda and J. R. Sanders, Eds., Washington, D.C.: National Council on Measurement in Education, 1979

Special Issue on Standard Setting, Journal of Educational Measurement, 1978 Vol. 15, No. 4. Contents: Standards and Criteria: Gene V Glass; Societal Standards: Nancy W. Burton; Responses: Michael Scriven, Ronald K. Hambleton, James H. Black, W. James Popham

Wilcox, R. R. and Harris, C. W. On Emrich's "An Evaluation Model for Mastery Testing", Journal of Educational Measurement, 1977, 14, 215-218

VII. Rasch Model

Applications of Latent Trait Models, (Special issue) Journal of Educational Measurement, Vol. 14, No. 2, 1977

Especially these articles:
Hambleton, R. K. and Cook, L. L. "Latent Trait Models and their use in the Analysis of Educational Test Data"

Wright, B. D. "Solving Measurement Problems with the Rasch Model"

Rentz, R. R. and Bashaw, W. L. "The National Reference Scale for Reading: An Application of the Rasch Model"

Wright, B. D. Sample-Free Test Calibration and Person Measurement. Proceedings of the 1969 Invitational Conference on Testing Problems. Princeton, N. J.: Educational Testing Service, 1968

Wright, B. D. and Stone, M. H. Best Test Design: Rasch Measurement. Chicago: MESA Press, 1979